High-dimension multi-label problems: convex or non convex relaxation?

Nicolas Papadakis, Romain Yildizoglu, Jean-François Aujol, Vicent Caselles

HIGH-DIMENSION MULTI-LABEL PROBLEMS: CONVEX OR NON CONVEX RELAXATION?

NICOLAS PAPADAKIS*, JEAN-FRANÇOIS AUJOL , VICENT CASELLES†, AND ROMAIN YILDIZOĞLU

Abstract. This paper is concerned with the problem of relaxing non convex functionals, used in image processing, into convex problems. We review most of the recently introduced relaxation methods, and we propose a new convex one based on a probabilistic approach, which has the advantages of being intuitive, flexible and involving an algorithm without inner loops. We investigate in detail the connections between the solutions of the relaxed functionals with a minimizer of the original one. Such connection is demonstrated only for non convex relaxation which turns out to be quite robust to initialization. As a case of study, we illustrate our theoretical analysis with numerical experiments, namely for the optical flow problem.

Key words. Multi-label problems, convex relaxation, segmentation, disparity and optical flow

AMS subject classifications. 15A15, 15A09, 15A23

1. Introduction. This paper is concerned with image processing problems whose solutions are computed as the minimizer of some functional (see e.g. [4]). The considered functionals have a data term which depends on the considered application, and a regularization term. Many image processing problems can be modeled as the minimizers of non convex functionals. This non convexity may arise from the physics of the problem as in speckle noise removal [3]. It may be due to the non locality of the functional [5]. It can also come from the problem itself as in image segmentation [23, 14, 6, 9, 21], optical flow computation [4, 25], ...

Non convexity of the functional to minimize may cause several issues. On one hand, the solution of the problem may not exist, and even when it exists may be non unique. On the other hand, numerical algorithms to compute the solution may get stuck into local minima, and the numerical solution may depend heavily on the initialization choice (in the case of iterative algorithms).

For all these reasons, it is a major improvement when a non convex problem can be turned into a convex one. Some steps in this direction have been done in the last past years [10, 24, 2, 28, 27, 25, 12, 30, 9, 19] within the mathematical image processing community. Convexification has been first used for image segmentation in [24], but it is now used for some other problems such as disparity computation [28], or higher dimensional problems like optical flow estimation [25]. The main idea of all these approaches is to introduce a new variable defined in a higher dimensional space that permits to write a convex relaxation of the original problem. If the original functional has some properties such as satisfying a layer cake formula, then a thresholding of the minimizer of the relaxed problem is a global minimizer of the original non convex one (see [10, 24] for seminal works in this direction). Of course, due to the curse of dimension, resorting to higher dimensional variables increases the computation time. This is the reason why convexification methods become more challenging when the

*N. Papadakis, J-F. Aujol and R. Yıldızoğlu are with Univ. Bordeaux, Institut de Mathmatiques de Bordeaux (IMB, UMR 5251), 351 Cours de la Libration F-33400 Talence, France (e-mail: jaujol@math.u-bordeaux1.fr, ryildi@gmail.com)

†V. Caselles is with the Departamento de Tecnologías de la Información y las Comunicaciones, Universitat Pompeu Fabra, Carrer de Roc Boronat 138, 08018 Barcelona, Spain. (e-mail: vicent.caselles@upf.edu)

dimension of the variables is larger than one. This is the case for the optical flow problem (as compared with the segmentation one [24]), and is the reason why we focus in this paper on optical flow computation.

This paper is inspired by the works we recalled in the previous paragraphs. It proposes new convexification choices, based on a probabilistic modeling and leading to new algorithms. We illustrate our approach on the optical flow problem. Our numerical experiments permit to derive some conclusions that justify the interest of the approach.

In all models proposed here, the regularization term is based on total variation. Such a term benefits from a layer cake formula thanks to the the coarea formula. This property plays a key role when one wants to relate the solution of the relaxed problem to the one of the original problem, as it will be detailed in the paper. Notice that some approaches such as [27, 30] propose to use a truncated version of total variation, as this was shown to be useful in image restoration (see [4]). This non convexity of the regularization term may lead to better preservation of discontinuities. However, as explained in the next paragraph, an important aspect of our work lies in the study of how to get back from the solution of the relaxed problem to the one of the original problem. The natural setting for such a study is to consider a classical total variation term for the regularization and not a non convex version of it.

The main contributions of the paper are the following:

- We provide the reader with a review of most of the recently introduced convexification methods in image processing. We discuss in particular the connections between these approaches, as well as some of their advantages and weaknesses.
- We introduce a general framework to relax image processing functionals into convex problems linked with the works of [28, 27, 12]. Our approach is based on a new probabilistic point of view which has the advantage of being flexible and intuitive. It also relies on an exact algorithm without any inner loops even for high dimensional problems.
- We propose two new relaxations for high dimensional problems: a non-convex relaxation in section 4.1.1 (which is separately convex in each variable), that represents a numerical and theoretical enhancement with respect to the model first presented in [25], and a convex relaxation in section 4.2. In order to make this paper as comprehensive as possible, we also review in section 4.1.2 the convexification method proposed in [30].
- We discuss in detail possible strategies to get back from a solution of the relaxed functionals to a minimizer of the original problem. As far as we know, this issue has not been investigated thoroughly in the literature yet. It is nevertheless a major problem, since computing a minimizer of the relaxed functional could be of no interest if it is not related to a minimizer of the original problem.
- We illustrate the different approaches presented in the paper, as well as the new ones, on different applications, including the optical flow problem. Since dimension is a major issue with convexification (as the dimension of the variables is increased when the functional is relaxed), we feel that optical flow is a good application to test the different frameworks (in view of the curse of dimension). We note that the different experiments focus on the ability of the proposed approach to recover global minima of the corresponding original problems.

The outline of the paper is as follows. We introduce the discrete framework we consider in Section 2. In particular, we detail the discretization choice, as well as the primal-dual algorithm that will be at the heart of all the numerical methods presented in the paper. For the sake of clarity, we first deal with the case of dimension 1 in Section 3. We first recall the framework introduced in [28], and then we introduce a new one based on a probabilistic modeling. In all these approaches, the solution of the original problem is computed from the one of the relaxed problem by a simple thresholding strategy. We conclude Section 3 by showing the connection of our probabilistic framework with the approach of Chambolle et al in [12].

We then concentrate on the dimension 2 case in section 4. We first recall the approach of [25] where the problem is relaxed independently for each variable (and this turns out to give a polyconvex energy). We also present the approach of [30] which proposes a way to convexify the data term of the functional; however, in this case, the relationship with the original problem is not clear, as will be illustrated in the numerical section. Then we,recall the framework of [9] that gives a general convexification method (which seems to be unrealistic to be used in practice for the optical flow problem due to the use of Dijkstra's algorithm to compute a complicated projection). Finally we introduce our probabilistic framework in the 2D setting and we show the equivalence between the two last relaxation methods. However, contrary to the 1D setting, there is no simple thresholding strategy to get back from the solution of the relaxed problem to the solution of the original one. This is the reason why we investigate in Section 5 the problem of computing a good solution of the original problem from the relaxed one. We then present extensive numerical examples in Section 6.

2. Preliminaries.

2.1. General problem. We focus our attention here on multi-label problems for image processing. Let Ω be the image domain, we assume Ω to be a non empty open bounded subset of $\mathbb{R}^2$ with Lipschitz boundary. We aim at estimating a set of N functions $u_i : \Omega \mapsto \Gamma_i$ (with $1 \leq i \leq N$) which take their values in a predefined discrete set containing M_i ordered elements: $\Gamma_i = \{u_0^i < \cdots < u_{M_i-1}^i\}$. This section is then dedicated to the minimization w.r.t $u = [u_1, \cdots, u_N]$ of the following class of functionals:

$$J_N(u) = \sum_{i=1}^{N} \int_\Omega |Du_i(x)| + \int_\Omega \rho(x, u(x))dx, \tag{2.1}$$

where ρ is a given positive data function and $\rho(x, u(x))$ represents the cost of assigning the values $u(x)$ to the pixel x. We only assume that ρ is a bounded function that can be non linear with respect to u. In the following, we will refer to the $1D$ case when $N = 1$.

The first terms $\int_\Omega |Du_i|$ contain the spatial regularization of the unknowns on the image domain. More precisely, they measure the integral of the perimeters of the level sets of u, assuming that u is a function of bounded variation (see [1] for more details). Such a term, introduced in image processing in the seminal work [29], is known as the total variation of u. This regularization is general and has been applied to a lot of image processing problems such as restoration [29], depth estimation [28], 3D reconstruction [20], or optical flow [32].

Let us recall some technical elements that will be used all along the paper.

2.2. Total variation.

Definition. The total variation of a function $u \in L^1(\Omega)$ is defined as $\int_\Omega |Du| := \sup\{\int_\Omega u \mathrm{div} z, z \in C_c^1(\Omega)^N, \| z \|_\infty \leq 1\}$, where $C_c^1(\Omega)$ denotes the space of C^1 functions with compact support in Ω (see [1]). Notice that in the case when u is a smooth function, then $\int_\Omega |Du| = \int_\Omega |\nabla u| \, dx$. We denote by $BV(\Omega, \mathbb{R})$ the space of functions with finite total variation in Ω.

Discretization schemes. The discretization used for the spatial gradient operator D_x and its adjoint D_x^* are the classical ones. We consider the discrete regular grid $x = (\mathrm{x}, \mathrm{y})$, $1 \leq \mathrm{x} \leq L_\mathrm{x}$, $1 \leq \mathrm{y} \leq L_\mathrm{y}$ representing the domain Ω. Looking at the discrete gradient operator as a vector of matrices $D_x = [D_\mathrm{x}, D_\mathrm{y}]^T$, the chosen discretizations should satisfy (to have a discrete Gauss-Green formula without boundary terms): $\langle D_\mathrm{x} u, z_1 \rangle + \langle D_\mathrm{y} u, z_2 \rangle = \langle u, D_\mathrm{x}^* z_1 + D_\mathrm{y}^* z_2) \rangle$. To that end, one can consider finite differences and take $D_x u(\mathrm{x}, \mathrm{y})$ with a forward scheme and $D_\mathrm{x}^* z$ with a backward one. The gradient with respect to the first dimension then reads

$$D_\mathrm{x} u(\mathrm{x}, \mathrm{y}) = \begin{cases} u(\mathrm{x}+1, \mathrm{y}) - u(\mathrm{x}, \mathrm{y}) & \text{if } 1 \leq \mathrm{x} < L_\mathrm{x}, \\ 0 & \text{if } \mathrm{x} = L_\mathrm{x}. \end{cases}$$

The corresponding divergence operator is given by $D_x^* \mathbf{z} = D_\mathrm{x}^* z_1 + D_\mathrm{y}^* z_2$, where the gradient over the first dimension is taken as:

$$D_\mathrm{x}^* z(\mathrm{x}, \mathrm{y}) = \begin{cases} z(\mathrm{x}, \mathrm{y}) & \text{if } \mathrm{x} = 1, \\ z(\mathrm{x}, \mathrm{y}) - z(\mathrm{x}-1, \mathrm{y}) & \text{if } 1 < \mathrm{x} < L_\mathrm{x}, \\ -z(\mathrm{x}-1, \mathrm{y}) & \text{if } \mathrm{x} = L_\mathrm{x}. \end{cases}$$

We consider analogous discretizations for $D_\mathrm{y} u$ and $D_\mathrm{y}^* z_2$. The discrete total variation of u can be defined as $\int_\Omega |D_x u| = \sum_{1 \leq \mathrm{x} \leq L_\mathrm{x}} \sum_{1 \leq \mathrm{y} \leq L_\mathrm{y}} |D_x u(\mathrm{x}, \mathrm{y})|$. Let us denote by $\mathcal{B}$ the set of vector fields $\mathbf{z} = (z_1, z_2)$, with z_i defined on $\Omega \times \Gamma_i$:

$$\mathcal{B} = \{\mathbf{z}, \, s.t. \, z_1^2(x, u_i) + z_2^2(x, u_i) \leq (u_i - u_{i-1})^2, \, \forall (x, u_i) \in \Omega \times \Gamma_i\}. \tag{2.2}$$

Then the total variation can be rewritten in its dual form $\int_\Omega |D_x u| = \max_{z \in \mathcal{B}} \langle u, D_x^* \mathbf{z} \rangle$.

2.3. Primal-dual algorithm.

Presentation. We recall the algorithm in [13]. U, Z are finite-dimensional vector spaces, we denote by $\langle ., . \rangle$ the standard inner products, $K : U \to Z$ is a linear operator, and $G : U \to \mathbb{R} \cup \{\infty\}$, $F^* : Z \to \mathbb{R} \cup \{\infty\}$ are convex functions. We want to solve

$$\min_{u \in U} \max_{z \in Z} \langle Ku, z \rangle + G(x) - F^*(y).$$

The algorithm

Algorithm 1 Primal-dual algorithm ([13])

$u^{k+1} = (I + \tau \partial G)^{-1}(u^k - \tau K^t z^k)$
$z^{k+1} = (I + \sigma \partial F^*)^{-1}(z^k + \sigma K(2u^{k+1} - u^k))$

with $(I + \tau \partial G)^{-1}(\hat{u}) := \arg\min_{u \in U} G(u) + \frac{1}{2\tau} \| u - \hat{u} \|^2$ and $\tau\sigma \| K \|^2 < 1$ converges to a saddle point in $O(\frac{1}{k})$. We have denoted by K^t the transpose of K. Notice that when K is the chosen discretized gradient operator, then $\|K\|^2 = 8$.

Proximal operator. If G is a convex proper lower semi continuous function, then $(I + \tau\partial G)^{-1}$ is the resolvent operator (which is a one to one mapping).

The proximity operator is defined by (see [22] for computation and details):

$$y = (I + \tau\partial G)^{-1}(x) = \text{prox}_h^G(x) = \underset{u}{\text{argmin}} \left\{ \frac{\|u - x\|^2}{2\tau} + G(u) \right\}. \tag{2.3}$$

We refer to [16] for examples of proximal operator computations. Notice that computing the proximal operator is itself a minimization problem. This computation amounts to computing a projection when G is the indicator of a closed convex set.

For instance, if $G = \chi_{z\in\mathcal{B}}$, where $\mathcal{B}$ is the set of vector fields $\mathbf{z} = (z_1, z_2)$ defined by Equation (2.2), then the resolvent operator $(I + \tau\partial G)^{-1}$ is (for any $\tau > 0$) the orthogonal projection:

$$P_{\mathcal{B}}(\mathbf{z}(x, u_i)) = \frac{(u_i - u_{i-1})\mathbf{z}(x, u_i)}{\max(u_i - u_{i-1}, ||\mathbf{z}(x, u_i)||)} \tag{2.4}$$

where $||\mathbf{z}(x, u_i)||^2 = z_1^2(x, u_i) + z_2^2(x, u_i)$.

3. Convexification of the $1D$ multi-label problem. This section is dedicated to the $1D$ case, where we only want to estimate 1 unknown for each pixel of an image. Concretely, we are interested in minimizing the functional:

$$J_1(u) = \int_\Omega |Du(x)| + \int_\Omega \rho(x, u(x))dx. \tag{3.1}$$

We assume that u takes values in $\Gamma = \{u_0 < \cdots < u_{M-1}\}$.

For the segmentation of a grayscale image $I : x \in \Omega \mapsto I(x) \in [0, 1]$ into M gray labels $\Gamma = \{0, 1/(M-1), \cdots 1\}$, the function ρ would read: $\rho(x, u(x)) = (I(x)-u(x))^2$. In the case of disparity estimation between two images I_1 and I_2, with disparity values in $\Gamma = \{0, 1, \cdots, M-1\}$, the cost would be: $\rho(x, u(x)) = (I_1(x) - I_2(x + u(x)))^2$.

3.1. Convexification of the $1D$ multi-label problem with upper level sets. We briefly recall here the convexification technique of Pock et al. [28, 27]. The idea is to write the non-linearities of the functional $J_1(u)$ in a convex way, by introducing an auxiliary variable $\phi : \Omega \times \Gamma \mapsto \{0, 1\}$ that represents the different values of u. The treatment here will be heuristic. Let

$$\phi(x, s) = H(u(x) - s), \tag{3.2}$$

where H is the Heaviside function ($H(r) = 1$ if $r \geq 0$, and 0 otherwise). The unknown u can then be recovered from ϕ by the layer cake formula as

$$u(x) = u_0 + \sum_{i=1}^{M-1} (u_i - u_{i-1})\phi(x, u_i). \tag{3.3}$$

Using the coarea formula, we find

$$\begin{aligned}\int_\Omega |Du(x)| &= \int_s \int_\Omega |D\mathbf{1}_{u\geq s}(x)|dxds = \sum_{i=1}^{M-1}(u_i - u_{i-1})\int_\Omega |D\mathbf{1}_{u\geq u_i}(x)|dx \\ &= \sum_{i=1}^{M-1}(u_i - u_{i-1})\int_\Omega |D_x\phi(x,u_i)|dx \\ \int_\Omega \rho(x,u(x)) &= \int_\Omega \sum_{i=0}^{M-1}\rho(x,u_i)\mathbf{1}_{u=u_i}(x)dx \\ &= \sum_{i=0}^{M-2}\int_\Omega \rho(x,u_i)(\phi(x,u_i)-\phi(x,u_{i+1}))dx + \int_\Omega \rho(x,u_{M-1})\phi(x,u_{M-1})dx.\end{aligned} \tag{3.4}$$

Hence we can rewrite functional (3.1) as a function of ϕ instead of u:

$$\begin{aligned}J_1(u) = &\sum_{i=1}^{M-1}(u_i - u_{i-1})\int_\Omega |D_x\phi(x,u_i)|dx + \sum_{i=0}^{M-2}\int_\Omega \rho(x,u_i)(\phi(x,u_i)-\phi(x,u_{i+1}))dx \\ &+ \int_\Omega \rho(x,u_{M-1})\phi(x,u_{M-1})dx = J(\phi),\end{aligned} \tag{3.5}$$

and we are led to consider the minimizing problem:

$$\min_{\phi\in\bar{\mathcal{A}}_1} J(\phi), \tag{3.6}$$

where

$$\bar{\mathcal{A}}_1 = \{\phi \in BV(\Omega\times\Gamma,\{0,1\}) \text{ such that } \ \phi(x,u_0)=1,\ \phi(x,u_i)\geq\phi(x,u_{i+1})\}. \tag{3.7}$$

Following [28], we have the following theorem.

THEOREM 3.1. *The layer cake formula (3.3) defines a bijection f between $\mathcal{A}_0$ and $BV(\Omega,\{0,1\})$ such that $J_1(f(\phi)) = J(\phi)$.*

Proof. We only need to verify that f is a bijection, the functional equality being given by (3.5). Let $\phi\in\mathcal{A}_0$, $x\in\Omega$, there exists j such that $\phi(x,u_i)=1$ for $i\leq j$ and $\phi(x,u_i)=0$ for $i>j$. Thus $\phi(x,.)$ is uniquely determined by this given j. We have $f(\phi)(x) = u_0 + \sum_{i=1}^{M-1}(u_i-u_{i-1})\phi(x,u_i) = u_j \in \Gamma$, and $f(\phi)$ is of bounded variation by (3.5), so f is well defined. Its inverse is $f^{(-1)}(u)(x,s) = H(u(x)-s)$. In other words, we identify a function with the characteristic function of its subgraph. □

Functional (3.5) is convex in ϕ. To find the global minimum of (3.1), one can find a ϕ minimizing (3.5), and then recover u from ϕ. Care must be taken to ensure that it is possible to compute u from ϕ. In particular, ϕ must be binary and decreasing with s. Unfortunately, the set of such functions ϕ is not convex. To recover convexity, the function ϕ should be allowed to take values on $[0,1]$. We introduce a convex set of admissible functions

$$\mathcal{A}_1 = \{\phi \in BV(\Omega\times\Gamma,[0,1]) \text{ such that } \phi(x,a)=1,\ \phi(x,u_i)\geq\phi(x,u_{i+1})\}, \tag{3.8}$$

and the convex problem

$$\min_{\phi\in\mathcal{A}_1} J(\phi). \tag{3.9}$$

Following [28], we have the following theorem.

THEOREM 3.2. *Let ϕ^* be a solution of (3.9), then for almost any threshold μ, $H(\phi^* - \mu)$ is a solution of (3.9) and of (3.6).*

Proof. Let us consider the solution ϕ^* of the relaxed problem (3.9). By the coarea formula we obtain:

$$\int_\Omega |D\phi^*| = \int_0^1 \left(\int_\Omega |DH(\phi^* - \mu)| \right) d\mu. \tag{3.10}$$

Furthermore

$$\begin{aligned} \int_\Omega \rho(x,u_i)\phi^*(x,u_i) &= \int_\Omega \rho(x,u_i) \int_0^1 H(\phi^*(x,u_i) - \mu) d\mu \\ &= \int_0^1 d\mu \int_\Omega \rho(x,u_i) H(\phi^*(x,u_i) - \mu). \end{aligned} \tag{3.11}$$

Using (3.5) we have that $J(\phi^*) = \int_0^1 J(H(\phi^* - \mu))d\mu$.

Let us now assume that $B := \{\mu \in [0,1]\colon\ H(\phi^* - \mu)$ is not a global solution of the non convex problem (3.6)$\}$ has a strictly positive measure. Then there exists $\phi' \in \mathcal{A}_1$ such that $J(\phi') < J(H(\phi^* - \mu))$ for every μ in B. This implies that $J(\phi') = \int_0^1 J(\phi')d\mu < \int_0^1 J(H(\phi^* - \mu))d\mu = J(\phi^*)$ which is impossible by definition of ϕ^*. ∎

Therefore, a global minimum of (3.1) can be recovered as a "cut" of ϕ. For almost any threshold $\mu \in [0,1]$, defining $\phi_\mu = \mathbf{1}_{\phi^* \geq \mu}$,

$$u(x) = u_0 + \sum_{i=1}^{M-1} \phi_\mu(x, u_i) \tag{3.12}$$

is a global minimum of (3.1).

3.1.1. Numerical optimization. The previous minimization problem (3.9) can be written as a primal-dual problem:

$$\min_{\phi \in \mathcal{A}_1} \max_{\mathbf{z} \in \mathcal{B}, z_3 \in \mathcal{C}} \int_\Omega \sum_{i=1}^{M-1} D_x\phi(x,u_i) \cdot \mathbf{z}(x,u_i)dx + \int_\Omega \sum_{i=0}^{M-1} D_u\phi(x,u_i) z_3(x,u_i)dx, \tag{3.13}$$

where the operator D_u and its adjoint D_u^* are defined as:

$$D_u\phi(x,u_i) = \begin{cases} \phi(x,u_{i+1}) - \phi(x,u_i) & \text{if } i < M-1, \\ -\phi(x,u_{M-1}) & \text{otherwise,} \end{cases} \tag{3.14}$$

$$D_u^* z(x,u_i) = \begin{cases} z(x,u_i) & \text{if } i = 0, \\ z(x,u_i) - z(x,u_{i-1}) & \text{if } 0 < i \leq M-1, \end{cases} \tag{3.15}$$

The orthogonal projection onto $\mathcal{B}$ denoted by $P_\mathcal{B}$ is recalled in Equation (2.4). The last dual variable $z_3 \in \Omega \times \Gamma$ is defined on the set

$$\mathcal{C} = \{z_3, \text{ s.t. } |z_3(x,u_i)| \leq \rho(x,u_i),\ \forall (x,u_i) \in \Omega \times \Gamma\}. \tag{3.16}$$

We consider the projection operator

$$P_\mathcal{C}(z_3(x,u_i)) = \begin{cases} \frac{z_3(x,u_i)}{|z_3(x,u_i)|}\rho(x,u_i) & \text{if } |z_3(x,u_i)| > \rho(x,u_i) \\ z_3(x,u_i) & \text{otherwise.} \end{cases}$$

We end up with the algorithm 2 to minimize the energy (3.5)

Algorithm 2 Minimizing the 1D energy (3.5)

Initialize $\phi^0(.,u_i) = \bar{\phi}^0(.,u_i) = \frac{u_{M-1}-u_i}{u_{M-1}-u_0}$, $z_i = 0$, choose $\sigma,\tau > 0$ such that $\sigma\tau < \frac{1}{8}$
while $||\phi^k - \phi^{k-1}|| > \epsilon$ **do**
$\mathbf{z}^{k+1} = P_{\mathcal{B}}\left(\mathbf{z}^k + \sigma D_x \bar{\phi}^k\right)$
$z_3^{k+1} = P_{\mathcal{C}}\left(z_3^k + \sigma D_u \bar{\phi}^k\right)$
$\phi^{k+1} = P_{\mathcal{A}_1}\left(\phi^k - \tau\left(D_x^* \mathbf{z}^{k+1} + D_u^* z_3^{k+1}\right)\right)$,
$\bar{\phi}^{k+1} = 2\phi^{k+1} - \phi^k$
end while

3.1.2. Important detail. The convex set $\mathcal{A}_1$ defined in Equation (3.8) involves the negativity of the quantities $\phi(x,u_{i+1}) - \phi(x,u_i)$. In the original paper [28], such property was not considered and the projection was just:

$$P_{\mathcal{A}_1}(\phi(x,u_i)) = \begin{cases} 1 & \text{if } i = 0, \\ \max(0,\min(1,\phi(x,u_i))) & \text{if } 0 < i < M-1, \\ 0 & \text{if } i = M-1. \end{cases} \tag{3.17}$$

This issue was fixed in [11], where an iterative algorithm inspired by Dijkstra's algorithm was introduced to define $P_{\mathcal{A}_1}$ and impose the constraint point-wise, for each $x \in \Omega$.

This is perhaps the main drawback of this convexification, as the process involves time consuming iterative projections (and the primal projection $P_{\mathcal{A}_1}$ is in theory only exact with an infinite number of Dijkstra iterations). For $1D$ problems, it has nevertheless been shown in [27] that such projection on decreasing constraint can be avoided, but the proof can not be extended to higher dimensions.

We now explain how a slight modification of the previous approach leads to simple projections for any $N \geq 1$. We found out that the proposed approach is connected with the recent one in [12], the one proposed here is based on an exact and simple dual projection.

3.2. A new convexification based on a probabilistic point of view. Recall that we are minimizing the functional

$$J_1(u) = \int_\Omega |Du(x)| + \int_\Omega \rho(x,u(x))dx. \tag{3.18}$$

The idea is to model the possible values of the data function ρ with

$$\min_{u\in\Gamma} \int_\Omega \rho(x,u(x))dx = \min_{w\in\mathcal{D}_1} \int_\Omega \sum_{i=0}^{M-1} \rho(x,u_i)w(x,u_i)dx,$$

where $w(x,u_i)$ represents the probability of assigning the label u_i to the pixel x. The equivalence with the original problem is obtained by considering a particular space for w that corresponds to a single assignation for each pixel $x \in \Omega$: there exists a unique $\tilde{s} \in \Gamma$ with $w(x,\tilde{s}) = 1$ and $w(x,s) = 0$ for $s \neq \tilde{s}$. Since probabilities with binary values are involved, this problem of assignation is not convex. The new variable w is then relaxed within the convex space

$$\mathcal{D}_1 = \left\{ w : \Omega \times \Gamma \mapsto [0,1], \text{ s.t. } \sum_i w(x,u_i) = 1,\ \forall x \in \Omega \right\}. \tag{3.19}$$

The Total Variation of u can be rewritten in terms of w as

$$\begin{aligned}\int_\Omega |Du| &= \int_\Omega \sum_{i=1}^{M-1} (u_i - u_{i-1}) |D_x H(u(x) - u_i)| ds \\ &= \int_\Omega \sum_{i=1}^{M-1} (u_i - u_{i-1}) \left| D_x \left(\sum_{j \geq i} w(x, u_j) \right) \right| dx,\end{aligned}$$

so that the functional finally is

$$J_w(w) = \int_\Omega \sum_{i=1}^{M-1} (u_i - u_{i-1}) \left| D_x \left(\sum_{j \geq i} w(x, u_j) \right) \right| dx + \int_\Omega \sum_{i=0}^{M-1} \rho(x, u_i) w(x, u_i) dx. \tag{3.20}$$

The problem of minimizing (3.20) w.r.t $w \in \mathcal{D}_1$ is convex and a global minimum can be estimated with the previously introduced optimization method involving a dual formulation of the TV term.

3.2.1. Numerical optimization. The dual formulation consists in rewriting the regularization term as

$$\begin{aligned}&\int_\Omega \sum_{i=1}^{M-1} (u_i - u_{i-1}) \left| D_x \left(\sum_{j \geq i} w(x, u_j) \right) \right| dx \\ =&\max_{\mathbf{z} \in \mathcal{B}} \int_\Omega \sum_{i=1}^{M-1} \left(\sum_{j \geq i} w(x, u_j) \right) D_x^* \mathbf{z}(x, u_i) dx,\end{aligned}$$

where the variable $\mathbf{z} = (z_1, z_2)$, with $z_k \in \Omega \times \Gamma$, is defined on the convex set given by Equation (2.2), that is,

$$\mathcal{B} = \{\mathbf{z},\ s.t.\ z_1^2(x, u_i) + z_2^2(x, u_i) \leq (u_i - u_{i-1})^2,\ \forall (x, u_i) \in \Omega \times \Gamma\}, \tag{3.21}$$

with the convention that $\mathbf{z}(., u_0) = 0$ since the level u_0 is useless for the TV energy of w as $\sum_{j \geq 0} w(x, u_j) = 1$, $\forall x \in \Omega$.

The projection operator $P_{\mathcal{B}}$ is given in Equation (2.4). The projection of w onto the convex set $\mathcal{D}_1$ (defined in Equation (3.19)) can finally be done point-wise for each pixel x by means of the projection of $w(x, .)$ onto a simplex of dimension M through the operator $P_{\mathcal{D}_1} : \mathbb{R}^{|\Omega| M} \to \mathcal{D}_1$ (see [15] for a fast implementation).

A sketch of the process for estimating w^* following the Primal-Dual approach of [13] and the discretizations given in section 2 is given in Algorithm 3. With respect to the previous section and the estimation of $\phi \in \mathcal{A}_1$ (defined in Equation (3.8)), the computational complexity is increased with the term $\sum_{l \geq i} w(x, u_l)$. However, the projection of the variable w is here exact [15], and not iterative as in the previous formulation. This makes the computational time for each iteration of both approaches equivalent, in case that we only do two iterations of Dijkstra's algorithm to project ϕ on $\mathcal{A}_1$ (and we note that 2 Dijkstra iterations are not sufficient in may cases). It is also relevant to note that preconditioning techniques [26] can be considered to speed-up the current process.

Algorithm 3 Minimizing the 1D energy (3.20)

Initialize $w^0 = \bar{w}^0 = \frac{1}{M}$, $\mathbf{z}^0 = 0$, choose $\sigma, \tau > 0$ such that $\sigma\tau < \frac{1}{8M}$
while $||w^k - w^{k-1}|| > \epsilon$ **do**

$$\mathbf{z}^{k+1}(., u_i) = P_{\mathcal{B}}\left(\mathbf{z}^k(., u_i) + \sigma D_x\left(\sum_{l\geq i} \bar{w}^k(., u_l)\right)\right), \qquad \forall u_i \in \Gamma$$

$$\tilde{w}(., u_i) = w^k(., u_i) - \tau\left(\rho(., u_i) + \sum_{l\leq i} D_x^* \mathbf{z}^{k+1}(., u_l)\right), \quad \forall u_i \in \Gamma$$

$$w^{k+1} = P_{\mathcal{D}_1}(\tilde{w}), \qquad \forall x \in \Omega$$

$$\bar{w}^{k+1} = 2w^{k+1} - w^k$$

end while

3.2.2. Equivalence with the original problem (3.18). Let us show that a solution of the original problem (i.e. a label assignation) can be obtained from the computed solution w^*. First of all, let us observe the relation with the approach in section 3, where the convexification is done with $\phi = H(u(x) - s)$, and the functional to minimize for $\phi \in \mathcal{A}_1$ is

$$J(\phi) = \sum_{i=1}^{M-1}(u_i - u_{i-1})\int_\Omega |D_x\phi(x, u_i)|dx + \sum_{i=0}^{M-2}\int_\Omega \rho(x, u_i)(\phi(x, u_i) - \phi(x, u_{i+1}))dx + \int_\Omega \rho(x, u_{M-1})(\phi(x, u_{M-1}))dx \tag{3.22}$$

with $\mathcal{A}_1 = \{\phi \in BV(\Omega \times \Gamma, [0, 1])$ such that $\phi(x, u_O) = 1,\ \phi(x, u_i) \geq \phi(x, u_{i+1})\}$.

PROPOSITION 3.3. *There exists a bijection f between the convex sets $\mathcal{A}_1$ and $\mathcal{D}_1$, with $f(\phi)(x, u_i) = \phi(x, u_i) - \phi(x, u_{i+1})$ and $f^{-1}(w)(x, u_i) = \sum_{j\geq i} w(x, u_j)$.*

Hence, setting $\phi(x, u_i) = \sum_{j\geq i} w(x, u_j)$ and observing that $w(x, u_i) = \phi(x, u_i) - \phi(x, u_{i+1})$, we have that

$$\min_{\phi\in\mathcal{A}_1} J(\phi) = \min_{w\in\mathcal{D}_1} J(f^{-1}(w)) = \min_{w\in\mathcal{D}_1} J_w(w). \tag{3.23}$$

As a consequence, if w^* is a global minimizer of the energy (3.20), then $\phi^* = f^{-1}(w^*)$ is a global minimizer of the energy (3.22).

We know from section 3 that the problem in ϕ is convex and that for almost any threshold μ then $H(\phi^* - \mu)$ is also a global solution. We can therefore build ϕ^* from w^* and threshold it to recover a global solution of the original problem.

3.3. Connections with the approach of Chambolle et al [12]. Notice that in [12], the authors have tackled the related problem of computing minimal partitions. Each label is represented by s_i and a cost σ_{ij} is the given distance between labels i and j. In our problem $\sigma_{ij} = |s_i - s_j|$. The problem is then to minimize w.r.t to a label map $\lambda : x \in \Omega \mapsto \lambda(x) = i \in [1, N]$, the energy:

$$J(\lambda) = \int_\Omega \rho(x, \lambda(x)) + \sum_i \sum_j \sigma_{ij}|\partial\Omega_i \cap \partial\Omega_j|,$$

where the partition of Ω in N areas is defined by

$$\Omega_i = \{x \in \Omega\, s.t.\, \lambda(x) = i\}.$$

The problem is relaxed by minimizing

$$\min_{w\in\mathcal{D}_1} \Psi(Dw) + \int_\Omega \rho(x,u_i)w(x,u_i)dx \tag{3.24}$$

with $\Psi(Dw) := \sup_{(q)\in K} \int_\Omega q_i.D_x w(x,u_i)$, $K = \{(q_i), |q_i - q_j| \leq |u_i - u_j|\}$. This is equivalent to our formulation. Indeed, we have

$$\begin{aligned}\sum_{i=1}^{M-1}(u_i - u_{i-1}) \int_\Omega |D\sum_{j\geq i} w_j| &= \sup_{|q_i|\leq u_i - u_{i-1}} \sum_{i=1}^{M-1} \int_\Omega q_i D \sum_{j\geq i} w_j \\ &= \sup_{|q_j|\leq u_i - u_{i-1}} \int_\Omega \sum_{j=1}^{N} Dw_j \left(\sum_{i\leq j} q_i\right).\end{aligned} \tag{3.25}$$

Setting $q'_j = \sum_{i\leq j} q_i$, we have $q_j = q'_{j+1} - q'_j$ and :

$$\sum_{i=1}^{M-1}(u_i - u_{i-1}) \int_\Omega |D\sum_{j\geq i} w_j| = \sup_{|q'_{j+1} - q'_j|\leq u_i - u_{i-1}} \sum_{j=1}^{M-1} \int_\Omega q'_j Dw_j, \tag{3.26}$$

which gives exactly the set $K = \{|q'_{j+1} - q'_j| \leq u_i - u_{i-1}\}$ which they prove to be the best relaxation for the minimal partition problem. As a consequence, our approach is similar to [12], but with a different probabilistic point of view.

4. Convexification of the $2D$ multi-label problem. We now look at the following class of functionals:

$$J_2(u,v) = \int_\Omega |Du(x)| + \int_\Omega |Dv(x)| + \int_\Omega \rho(x,u(x),v(x))dx, \tag{4.1}$$

defined for the variables $u(x)$ and $v(x)$ taking their values in the discrete sets $\Gamma^u = \{u_0 < \cdots u_{M-1}\}$ and $\Gamma^v = \{v_0 < \cdots v_{M-1}\}$. For sake of clarity we will now consider that the discretization step of Γ^u and Γ^v is uniform, so that $u_i - u_{i-1} = \Delta_u$, and $v_i - v_{i-1} = \Delta_v$ for $i = 1 \cdots M-1$.

Let us give an example of the data term ρ for the problem of optical flow estimation between two images. In this case, we seek the $2D$ vector field corresponding to the flow $w(x) = (u(x), v(x))$ going from image I_1 to image I_2. The field $w(x) = (u(x), v(x))$ represents the displacement field of pixels x of image I_1. The cost function ρ can then be defined as

$$\rho(x,u(x),v(x)) = \min(\kappa, |I_1(x) - I_2(x + w(x))|), \tag{4.2}$$

where $\kappa > 0$ is used to threshold the data term and deal with local artifacts. We just set $\rho(x,.,.) = \kappa$ when $x + w(x)$ is outside Ω.

4.1. Independent relaxation. In this part, we consider the natural extension of the previous section that consists in introducing an auxiliary function for each variable to estimate. Following the probability formulation of section 3.2, we now have two functions w_u and w_v that belong to the convex set of admissible functions $\mathcal{D}_1$, defined in (3.19). The regularization terms of energy (4.1) can thus be written as

$$\begin{aligned}\int_\Omega |Du| &= \int_\Omega \sum_{i=1}^{M-1} (u_i - u_{i-1}) |D_x H(u(x) - u_i)| ds \\ &= \Delta_u \int_\Omega \sum_{i=1}^{M-1} \left| D_x \left(\sum_{j \geq i} w_u(x, u_j) \right) \right| dx \\ \int_\Omega |Dv| &= \Delta_v \int_\Omega \sum_{i=1}^{M-1} \left| D_x \left(\sum_{j \geq i} w_v(x, v_j) \right) \right| dx\end{aligned} \tag{4.3}$$

Let us now detail how to treat the data term.

4.1.1. Non convex data term. Using the previously introduced notations, the data term reads

$$\begin{aligned}\int_\Omega \rho(x, u(x), v(x)) dx &= \int_\Omega \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} \rho(x, u_i, v_j) \mathbf{1}_{u=u_i}(x) \mathbf{1}_{v=v_j}(x) dx \\ &= \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} \int_\Omega \rho(x, u_i, v_j) w_u(x, u_i) w_v(x, v_j),\end{aligned} \tag{4.4}$$

and we end up with the following problem

$$\min_{(w_u, w_v) \in \mathcal{D}_1 \times \mathcal{D}_1} J(w_u, w_v), \tag{4.5}$$

with

$$\begin{aligned}J(w_u, w_v) = &\sum_{i=0}^{M-1} \sum_{j=0}^{M-1} \int_\Omega \rho(x, u_i, v_j) w_u(x, u_i) w_v(x, v_j) \\ &+ \Delta_u \int_\Omega \sum_{i=1}^{M-1} \left| D_x \left(\sum_{j \geq i} w_u(x, u_j) \right) \right| dx + \Delta_v \int_\Omega \sum_{i=1}^{M-1} \left| D_x \left(\sum_{j \geq i} w_v(x, v_j) \right) \right| dx.\end{aligned} \tag{4.6}$$

The data term of this energy is not jointly convex in the variables w_u and w_v, so that we will only be able to reach a local minima of the problem (4.5). Nevertheless, we are able to recover some interesting properties on the obtained local minima of the non convex functional. Using the upper level set formulation and extending Proposition 3.3 to the 2D case, the functional can be equivalently reformulated as

$$\begin{aligned}J(\phi_u, \phi_v) = &\sum_{i=1}^{M-1} \int_\Omega |D_x \phi_u(x, u_i)| dx + \sum_{j=1}^{M-1} \int_\Omega |D_x \phi_v(x, v_j)| dx \\ &+ \sum_{i=0}^{M-2} \sum_{j=0}^{M-2} \int_\Omega \rho(x, u_i, v_j)(\phi_u(x, u_i) - \phi_u(x, u_{i+1}))(\phi_v(x, v_i) - \phi_v(x, v_{j+1})) dx \\ &+ \sum_{j=1}^{M-2} \int_\Omega \rho(x, u_{M-1}, v_j) \phi_u(x, u_{M-1})(\phi_v(x, v_i) - \phi_v(x, v_{j+1})) dx \\ &+ \sum_{i=1}^{M-2} \int_\Omega \rho(x, u_i, v_{M-1})(\phi_u(x, u_i) - \phi_u(x, u_{i+1})) \phi_v(x, v_{M-1}) dx \\ &+ \int_\Omega \rho(x, u_{M-1}, v_{M-1}) \phi_u(x, u_{M-1}) \phi_v(x, v_{M-1}) dx\end{aligned} \tag{4.7}$$

with $\phi_u \in \mathcal{A}_u$ and $\phi_v \in \mathcal{A}_v$ defined in

$$\begin{aligned}\mathcal{A}_u &= \{\phi_u \in BV(\Omega \times \Gamma^u, \{0,1\}) \text{ such that } \ \phi_u(x,u_0)=1,\ \phi_u(x,u_i) \geq \phi_u(x,u_{i+1})\},\\ \mathcal{A}_v &= \{\phi_v \in BV(\Omega \times \Gamma^v, \{0,1\}) \text{ such that } \ \phi_v(x,v_0)=1,\ \phi_v(x,v_j) \geq \phi_v(x,v_{j+1})\}.\end{aligned} \tag{4.8}$$

The functional J is convex in ϕ_u and in ϕ_v but not in (ϕ_u, ϕ_v). We have the following properties.

PROPOSITION 4.1. *Let $(\phi_u^*, \phi_v^*) \in \mathcal{A}_u \times \mathcal{A}_v$ be a local minimum of J, then, for almost any $\mu \in [0,1]$, $J(H(\phi_u^* - \mu), \phi_v^*) = J(\phi_u^*, H(\phi_v^* - \mu)) = J(\phi_u^*, \phi_v^*)$.*

Proof. We know from the co-area theorem that $J(\phi_u, \phi_v) = \int_0^1 J(H(\phi_u - \mu), \phi_v) d\mu$. By contradiction, assume that there exists μ such that $J(H(\phi_u^* - \mu), \phi_v^*) < J(\phi_u^*, \phi_v^*)$. As (ϕ_u^*, ϕ_v^*) is a local minimum, for any small $\epsilon > 0$, we have:

$$J(\phi_u^*, \phi_v^*) \leq J((1-\epsilon)\phi_u^* + \epsilon H(\phi_u^* - \mu), \phi_v^*).$$

From the convexity of J in its first variable ϕ_u, we have $\forall \epsilon \in (0,1)$:

$$\begin{aligned} J(\phi_u^*, \phi_v^*) &\leq J((1-\epsilon)\phi_u^* + \epsilon H(\phi_u^* - \mu), \phi_v^*) \\ &\leq (1-\epsilon) J(\phi_u^*, \phi_v^*) + \epsilon J(H(\phi_u^* - \mu), \phi_v^*) \\ &< J(\phi_u^*, \phi_v^*), \end{aligned} \tag{4.9}$$

which leads to a contradiction. Thus we have $J(H(\phi_u^* - \mu), \phi_v^*) \geq J(\phi_u^*, \phi_v^*)$ for any $\mu \in [0,1]$. Since the co-area formula holds, this implies that $J(H(\phi_u^* - \mu), \phi_v^*) = J(\phi_u^*, \phi_v^*)$ for any $\mu \in [0,1]$ (first we deduce this for almost any $\mu \in [0,1]$ and then for any $\mu \in [0,1]$ by the lower semi-continuity of the energy). ∎

PROPOSITION 4.2. *Let $(\phi_u^*, \phi_v^*) \in \mathcal{A}_u \times \mathcal{A}_v$ be a local minima of J, then for almost any $(\mu, \nu) \in [0,1]^2$, $J(H(\phi_u^* - \mu), H(\phi_v^* - \nu)) = J(\phi_u^*, \phi_v^*)$.*

Proof. Assume that there exist μ, ν such that $J(H(\phi_u^* - \mu), H(\phi_v^* - \nu)) < J(\phi_u^*, \phi_v^*)$. From the previous proposition we know that $J(H(\phi_u^* - \mu), \phi_v^*) = J(\phi_u^*, \phi_v^*)$. With the convexity of J in its second variable ϕ_v, we have $\forall \epsilon \in (0,1)$:

$$\begin{aligned} &\underbrace{J(H(\phi_u^* - \mu), (1-\epsilon)\phi_v^* + \epsilon H(\phi_v^* - \nu))}_{A} \\ \leq &(1-\epsilon) \underbrace{J(H(\phi_u^* - \mu), \phi_v^*)}_{= J(\phi_u^*, \phi_v^*)} + \epsilon \underbrace{J(H(\phi_u^* - \mu), H(\phi_v^* - \nu))}_{< J(\phi_u^*, \phi_v^*)} \\ < &J(\phi_u^*, \phi_v^*). \end{aligned} \tag{4.10}$$

With the convexity of J in its first variable, we can also notice that

$$\begin{aligned} &J((1-\epsilon)\phi_u^* + \epsilon H(\phi_u^* - \mu), (1-\epsilon)\phi_v^* + \epsilon H(\phi_v^* - \nu)) \\ \leq &(1-\epsilon) \underbrace{J(\phi_u^*, (1-\epsilon)\phi_v^* + \epsilon H(\phi_v^* - \nu))}_{= J(\phi_u^*, \phi_v^*)} + \epsilon \underbrace{J(H(\phi_u^* - \mu), (1-\epsilon)\phi_v^* + \epsilon H(\phi_v^* - \nu))}_{= A < J(\phi_u^*, \phi_v^*)} \\ < &J(\phi_u^*, \phi_v^*). \end{aligned} \tag{4.11}$$

However, if $\epsilon \to 0$ (with $\epsilon > 0$) then $(1-\epsilon)\phi_u^* + \epsilon H(\phi_u^* - \mu), (1-\epsilon)\phi_v^* + \epsilon H(\phi_v^* - \nu)$ belongs to the neighborhood of the local minima (ϕ_u^*, ϕ_v^*), so we necessarily have:

$$\lim_{\epsilon \to 0} J((1-\epsilon)\phi_u^* + \epsilon H(\phi_u^* - \mu), (1-\epsilon)\phi_v^* + \epsilon H(\phi_v^* - \nu)) \geq J(\phi_u^*, \phi_v^*),$$

which leads to a contradiction. ∎

Hence, we know that any thresholding of a local minima will give us a solution of the original problem with the same energy. To obtain a local minima, we consider the following approaches.

Alternate minimization. Assuming that w_v (resp. w_u) is known, the functional can be minimized alternatively with respect to w_u (resp. w_v) until convergence [25] . In both cases, we can linearize with respect to the optimized variable and recover the original $1D$ problem. For instance, when optimizing with respect to ϕ, we have the problem

$$\min_{w_u} \Delta_u \int_\Omega \sum_{i=1}^{M-1} \left| D_x \left(\sum_{j\geq i} w_u(x,u_j) \right) \right| dx + \sum_{i=0}^{M-1} \int_\Omega \rho_u(x,u_i) w_u(x,u_i), \tag{4.12}$$

where the data value ρ_u is obtain by integration over the dimension associated to v:

$$\rho_u(x,u_i) = \sum_{j=0}^{M-1} \rho(x,u_i,v_j) w_v(x,v_j). \tag{4.13}$$

The theoretical optimization tools of section 3.1.1 can be used to minimize (4.12). At convergence of the process, the obtained local minima (w_u^*, w_v^*) can be used to build $\phi_u = f(w_u^*)$ and $\phi_v = f(w_v^*)$, with the function f defined in Proposition 3.3.

By thresholding such functions, we can produce an admissible function of the original non convex problem : $u^*(x) = u_0 + \Delta_u \sum_{i=1}^{M-1} \phi_u(x,u_i,)$ and $v^*(x) = v_0 + \Delta_v \sum_{i=1}^{M-1} \phi_v(x,v_i)$. However, we here have no information on the obtained couple $(u^*(x), v^*(x))$, which has no reason to be a local extremum.

Joint minimization. In order to minimize properly the non convex functional (4.6), the joint descent direction can be considered for the couple (w_u, w_v). Hence, the optimal relations given by the partial derivatives of (4.6) w.r.t (w_u, w_v) read:

$$\begin{aligned}
w_u^{k+1}(.,u_i) &= w_u^k(.,u_i) - \tau \left(\sum_{l\leq i} D_x^* \mathbf{z}_u^{k+1}(.,u_l) + \sum_{j=0}^{M-1} \rho(.,u_i,v_j) w_v^{k+1}(.,v_j) \right) \\
w_v^{k+1}(.,v_j) &= w_v^k(.,v_j) - \tau \left(\sum_{l\leq j} D_x^* \mathbf{z}_v^{k+1}(.,v_l) + \sum_{i=0}^{M-1} \rho(.,u_i,v_j) w_u^{k+1}(.,u_i) \right)
\end{aligned}$$

The solution (w_u^{k+1}, w_v^{k+1}) can be obtained by solving for each point $x \in \Omega$:

$$\begin{bmatrix} w_u^{k+1}(x,u_1) \\ \vdots \\ w_u^{k+1}(x,u_M) \\ w_v^{k+1}(x,v_1) \\ \vdots \\ w_v^{k+1}(x,v_M) \end{bmatrix} = \underbrace{\begin{bmatrix} Id & \tau\rho(x,.,.) \\ \tau\rho^T(x,.,.) & Id \end{bmatrix}^{-1}}_{A} \begin{bmatrix} w_u^k(x,u_1) - \tau \sum_{l\leq 1} D_x^* \mathbf{z}_u^{k+1}(x,u_l) \\ \vdots \\ w_u^k(x,u_M) - \tau \sum_{l\leq M} D_x^* \mathbf{z}_u^{k+1}(x,u_l) \\ w_v^k(x,v_1) - \tau \sum_{l\leq 1} D_x^* \mathbf{z}_v^{k+1}(x,v_l) \\ \vdots \\ w_v^k(x,v_M) - \tau \sum_{l\leq M} D_x^* \mathbf{z}_v^{k+1}(x,v_l) \end{bmatrix}$$

where the constant matrix A of size $4M^2$ is constant and can be inverted beforehand. This approach therefore requires a lot of memory, but we will see in the experimental section that it produces results of the same quality (or even better accuracy) than the convex relaxations. The full process is given in algorithm 5.

Algorithm 4 Alternate minimization of the 2D non-convex energy (4.6)

Initialize $w_u^0 = \bar{w}_u^0 = \frac{1}{M}$, $w_v^0 = \bar{w}_v^0 = \frac{1}{M}$,$\mathbf{z}_u^0 = \mathbf{z}_v^0 = 0$, $\sigma, \tau > 0$ such that $\sigma\tau < \frac{1}{8M}$
while $||w_u^l - w_u^{l-1}|| + ||w_v^l - w_v^{l-1}|| > \epsilon$ **do**
  Compute $\rho_u(x, u_i) = \sum_{j=0}^{M-1} \rho(x, u_i, v_j) w_v^l(x, v_j)$
  Set $w_u^0 = w_u^{l-1}$
  while $||w_u^k - w_u^{k-1}|| > \epsilon$ **do**

$$\mathbf{z}_u^{k+1}(., u_i) = P_{\mathcal{B}}\left(\mathbf{z}_u^k(., u_i) + \sigma D_x\left(\sum_{l \geq i} \bar{w}_u^k(., u_l)\right)\right), \qquad \forall u_i \in \Gamma^u$$

$$\tilde{w}_u(., u_i) = w_u^k(., u_i) - \tau\left(\rho_u(., u_i) + \sum_{l \leq i} D_x^* \mathbf{z}_u^{k+1}(., u_l)\right), \quad \forall u_i \in \Gamma^u$$

$$w_u^{k+1} = P_{\mathcal{D}_1}(\tilde{w}_u), \qquad \forall x \in \Omega$$

$$\bar{w}_u^{k+1} = 2w_u^{k+1} - w_u^k$$

  end while
  Set $w_u^{l+1} = w_u^{k+1}$
  Compute $\rho_v(x, v_j) = \sum_{i=0}^{M-1} \rho(x, u_i, v_j) w_u^{l+1}(x, u_i)$
  Set $w_v^0 = w_v^{l-1}$
  while $||w_v^k - w_v^{k-1}|| > \epsilon$ **do**

$$\mathbf{z}_v^{k+1}(., v_j) = P_{\mathcal{B}}\left(\mathbf{z}_v^k(., v_j) + \sigma D_x\left(\sum_{l \geq j} \bar{w}_v^k(., v_l)\right)\right), \qquad \forall v_j \in \Gamma^v$$

$$\tilde{w}_v(., v_j) = w_v^k(., v_j) - \tau\left(\rho_v(., v_j) + \sum_{l \leq j} D_x^* \mathbf{z}_v^{k+1}(., v_l)\right), \quad \forall v_i \in \Gamma^v$$

$$w_v^{k+1} = P_{\mathcal{D}_1}(\tilde{w}_v), \qquad \forall x \in \Omega$$

$$\bar{w}_v^{k+1} = 2w_v^{k+1} - w_v^k,$$

  end while
  Set $w_v^{l+1} = w_v^{k+1}$
end while

Algorithm 5 Minimizing the 2D non-convex energy (4.6)

Initialize $w_u^0 = \bar{w}_u^0 = \frac{1}{M}$, $w_v^0 = \bar{w}_v^0 = \frac{1}{M}$, $\mathbf{z}_u^0 = \mathbf{z}_v^0 = 0$, choose $\sigma, \tau > 0$ such that $\sigma\tau < \frac{1}{8M}$
while $||w_u^k - w_u^{k-1}|| + ||w_v^k - w_v^{k-1}|| > \epsilon$ **do**

$$\mathbf{z}_u^{k+1}(., u_i) = P_{\mathcal{B}}\left(\mathbf{z}_u^k(., u_i) + \sigma D_x\left(\sum_{l \geq i} \bar{w}_u^k(., u_l)\right)\right), \quad \forall u_i \in \Gamma^u$$

$$\mathbf{z}_v^{k+1}(., v_j) = P_{\mathcal{B}}\left(\mathbf{z}_v^k(., v_j) + \sigma D_x\left(\sum_{l \geq j} \bar{w}_v^k(., v_l)\right)\right), \quad \forall v_j \in \Gamma^v$$

$$b_u(., u_i) = w_u^k(., u_i) - \tau \sum_{l \leq i} D_x^* \mathbf{z}_u^{k+1}(., u_l), \qquad \forall u_i \in \Gamma^u$$

$$b_v(., v_j) = w_v^k(., v_j) - \tau \sum_{l \leq j} D_x^* \mathbf{z}_v^{k+1}(., v_l), \qquad \forall v_j \in \Gamma^v$$

$$\begin{bmatrix} \tilde{w}_u(x, .) \\ \tilde{w}_v(x, .) \end{bmatrix} = A \begin{bmatrix} b_u(x, .) \\ b_v(x, .) \end{bmatrix}, \qquad \forall x \in \Omega$$

$$w_u^{k+1} = P_{\mathcal{D}_1}(\tilde{w}_u), \qquad \forall x \in \Omega$$

$$\bar{w}_u^{k+1} = 2w_u^{k+1} - w_u^k$$

$$w_v^{k+1} = P_{\mathcal{D}_1}(\tilde{w}_v), \qquad \forall x \in \Omega$$

$$\bar{w}_v^{k+1} = 2w_v^{k+1} - w_v^k$$

end while

Remark. When looking at the continuous formulation (Δ_u and Δ_v tending to 0) within the upper level sets framework, the data term reads:

$$\int_{\Gamma^u}\int_{\Gamma^v}\int_{\Omega}\rho(x,s,t)\partial_s\phi_u(x,s)\partial_t\phi_v(x,t)dsdtdx.$$

It has been noticed in [25] that this term is polyconvex in the sense of [8] as it involves the determinant of the Jacobian Matrix in (s,t):

$$\begin{pmatrix}\partial_s\phi & \partial_t\phi\\ \partial_s\psi & \partial_t\psi\end{pmatrix}=\begin{pmatrix}\partial_s\phi & 0\\ 0 & \partial_t\psi\end{pmatrix}.$$

Polyconvex functionals are quasiconvex. As quasiconvexity is the right extension of the notion of convexity for vector valued functions, it guarantees, under assumptions, the existence of minimizers and the well-posedness (in a certain sense) of the energies. Since the probability formulation shares the properties of the upper level sets one, it may explain the good behavior of the non convex approach on real experiments that will be detailed in the application section.

4.1.2. Convexification of the data term. To tackle properly the whole energy, some convex relaxations of the data term have been proposed in [17, 30]. The idea of [30] is to compute the biconjugate of the data term with the Legendre-Fenchel transform in order to obtain a tight relaxation. This leads to reformulate the data term for binary variables (w_u^b, w_v^b) in the following convex way:

$$\begin{aligned}&\sum_{i=0}^{M-1}\sum_{j=0}^{M-1}\int_{\Omega}\rho(x,u_i,v_j)w_u^b(x,u_i)w_v^b(x,v_j)\\ =&\max_{(q_u,q_v)\in\mathcal{Q}}\sum_{i=0}^{M-1}\int_{\Omega}\left(w_u^b(x,u_i)q_u(x,u_i)+w_v^b(x,v_i)q_v(x,v_i)\right),\end{aligned}$$

where two other auxiliary variables $q_u(x,u_i)$ and $q_v(x,v_j)$ have been introduced. They are defined in the convex set:

$$\begin{aligned}\mathcal{Q}=&\{q_u:\Omega\times\Gamma_u\mapsto\mathbb{R},\ q_v:\Omega\times\Gamma_v\mapsto\mathbb{R},\\ &\text{s.t. } q_u(x,u_i)+q_v(x,v_j)\leq\rho(x,u_i,v_j)\,\forall x\in\Omega,\ 0\leq i,j\leq M-1\}.\end{aligned}$$

Notice that for non binary variables (w_u, w_v), we will have:

$$\begin{aligned}&\sum_{i=0}^{M-1}\sum_{j=0}^{M-1}\int_{\Omega}\rho(x,u_i,v_j)w_u(x,u_i)w_v(x,v_j)\\ \geq&\max_{(q_u,q_v)\in\mathcal{Q}}\sum_{i=0}^{M-1}\int_{\Omega}\left(w_u(x,u_i)q_u(x,u_i)+w_v(x,v_i)q_v(x,v_i)\right).\end{aligned}\tag{4.14}$$

Slightly modifying the approach of [30], we end up with the following energy to minimize w.r.t w_u and w_v:

$$\begin{aligned}J_C(w_u,w_v)=&\max_{(q_u,q_v)\in\mathcal{Q}}\sum_{i=0}^{M-1}\int_{\Omega}\left(w_u(x,u_i)q_u(x,u_i)+w_v(x,v_i)q_v(x,v_i)\right)\\ &+\Delta_u\int_{\Omega}\sum_{i=1}^{M-1}\left|D_x\left(\sum_{j\geq i}w_u(x,u_j)\right)\right|dx+\Delta_v\int_{\Omega}\sum_{i=1}^{M-1}\left|D_x\left(\sum_{j\geq i}w_v(x,v_j)\right)\right|dx.\end{aligned}\tag{4.15}$$

As the orthogonal projection on $\mathcal{Q}$ can not be written explicitly, the corresponding proximal operator needed by the primal-dual algorithm of [13] can not be formulated directly for the new variables (q_u, q_v). One solution is therefore to do inner loops in order to approximate the projection onto $\mathcal{Q}$. The authors of [30] chose to consider a Lagrange multiplier, by introducing a new variable $\lambda(x, u_i, v_j) \in \Omega \times \Gamma^u \times \Gamma^v$ to enforce the constraint $q_u(x, u_i) + q_v(x, v_j) \leq \rho(x, u_i, v_j)$. The projection onto $\mathcal{Q}$ is therefore done by solving

$$\begin{aligned} P_{\mathcal{Q}}(\tilde{q}_u, \tilde{q}_v) = \min_{(q_u,q_v)} \max_{\lambda \geq 0} & ||q_u - \tilde{q}_u||^2 + ||q_v - \tilde{q}_v||^2 \\ & + \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} \int_\Omega \left(q_u(x, u_i) + q_v(x, v_j) - \rho(x, u_i, v_j)\right) \lambda(x, u_i, v_j), \end{aligned} \tag{4.16}$$

with an Uzawa's algorithm. Such projection then requires additional parameters for monitoring the inner loops. In our experiments, we realized a fixed number of 50 iterations in our implementation to approximate the projection on $\mathcal{Q}$, that turns out to be a good compromise between computational cost and quality of the results. The process converges perfectly with such parametrization but the complexity of the projection is not negligible as it consists in iterations in the domain of $\lambda \in \mathbb{R}^{|\Omega| M^2}$. The final process is illustrated in Algorithm 6.

With this new convexification of the data term and contrary to the previous non convex approach, the equivalence with the upper level set formulation does no longer hold, and the layer cake formula can no longer be used. To be more precise, let us rewrite the new data term within the upper level set formulation:

$$\begin{aligned} & D(w_u, w_v) \\ = & \max_{(q_u,q_v)\in\mathcal{Q}} \sum_{i=0}^{M-1} \int_\Omega (w_u(x, u_i) q_u(x, u_i) + w_v(x, v_i) q_v(x, v_i)) \\ = & \max_{(q_u,q_v)\in\mathcal{Q}} \sum_{i=0}^{M-2} \int_\Omega (\phi_u(x, u_i) - \phi_u(x, u_{i+1}) q_u(x, u_i) + (\phi_v(x, v_i) - \phi_v(x, v_{i+1}) q_v(x, v_i)) \\ & + \int_\Omega \phi_u(x, u_{M-1}) q_u(x, u_{M-1}) + \int_\Omega \phi_v(x, v_{M-1}) q_v(x, v_{M-1}) \\ & = \max_{(q_u,q_v)\in\mathcal{Q}} \tilde{D}(\phi_u, \phi_v, q_u, q_v) \\ & = \tilde{D}(\phi_u, \phi_v, q_u^*, q_v^*) \\ & = D(\phi_u, \phi_v), \end{aligned}$$

where the couple $(q_u^*, q_v^*) \in \mathcal{Q}$ maximizes the previous relation for (ϕ_u, ϕ_v) given. One can therefore observe that

$$\begin{aligned} D(\phi_u, \phi_v) &= \int_0^1 \tilde{D}(H(\phi_u - \mu), H(\phi_v - \mu), q_u^*, q_v^*) d\mu \\ &\leq \int_0^1 \max_{(q_u,q_v)\in\mathcal{Q}} \tilde{D}(H(\phi_u - \mu), H(\phi_v - \mu), q_u, q_v) d\mu \\ &\leq \int_0^1 D(H(\phi_u - \mu), H(\phi_v - \mu)) d\mu. \end{aligned}$$

From relation (4.14), we thus have:

$$J_C(w_u, w_v) \leq J(w_u, w_v),$$

where J_C and J respectively refer to expressions (4.6) and (4.15). We can only prove that both functionals are equivalent for binary arguments. As a consequence, if the computed global minimum of J_C is not binary, we have no guarantee that a thresholding operation will be able to recover a minimum (neither local or global) for both the relaxed non convex problem (4.6) and the original problem (4.1). This is the reason why the authors of [30] proposed to build an admissible function of the original problem by keeping for each pixel x, the couple of labels (u_i, v_j) that maximize $w_u^*(x, u_i) + w_v^*(x, v_i)$.

Algorithm 6 Minimizing the 2D convex energy (4.15)

Initialize $w_u^0 = \bar{w}_u^0 = \frac{1}{M}$, $w_v^0 = \bar{w}_v^0 = \frac{1}{M}$, $\mathbf{z}_u^0 = \mathbf{z}_v^0 = q_u^0 = q_v^0 = 0$, choose $\sigma, \tau > 0$ such that $\sigma\tau < \frac{1}{8M^2}$

while $||w_u^k - w_u^{k-1}|| + ||w_v^k - w_v^{k-1}|| > \epsilon$ **do**

$$\mathbf{z}_u^{k+1}(., u_i) = P_{\mathcal{B}}\left(\mathbf{z}_u^k(., u_i) + \sigma D_x\left(\textstyle\sum_{l\geq i} \bar{w}_u^k(., u_l)\right)\right), \qquad \forall u_i \in \Gamma^u$$

$$\mathbf{z}_v^{k+1}(., v_j) = P_{\mathcal{B}}\left(\mathbf{z}_v^k(., v_j) + \sigma D_x\left(\textstyle\sum_{l\geq j} \bar{w}_v^k(., v_l)\right)\right), \qquad \forall v_j \in \Gamma^v$$

$$\tilde{q}_u(., u_i) = q_u^k(., u_i) + \sigma(\bar{w}_u^k(., u_i) + \textstyle\sum_{j=1}^M \lambda(., u_i, v_j)), \qquad \forall u_i \in \Gamma^u$$

$$\tilde{q}_v(., v_j) = q_v^k(., v_j) + \sigma(\bar{w}_v^k(., v_j) + \textstyle\sum_{i=1}^M \lambda(., u_i, v_j)), \qquad \forall v_j \in \Gamma^v$$

$$[q_u^{k+1}, q_v^{k+1}] = P_{\mathcal{Q}}([\tilde{q}_u, \tilde{q}_v]), \qquad \forall x \in \Omega$$

$$\tilde{w}_u(., u_i) = w_u^k(., u_i) - \tau\left(q_u^{k+1}(., u_i) + \textstyle\sum_{l\leq i} D_x^* \mathbf{z}_u^{k+1}(., u_l)\right), \qquad \forall u_i \in \Gamma^u$$

$$w_u^{k+1} = P_{\mathcal{D}}\left(\tilde{w}_u\right), \qquad \forall x \in \Omega$$

$$\bar{w}_u^{k+1} = 2w_u^{k+1} - \bar{w}_u^k$$

$$\tilde{w}_v(., v_j) = w_v^k(., v_j) - \tau\left(q_v^{k+1}(., v_j) + \textstyle\sum_{l\leq j} D_x^* \mathbf{z}_v^{k+1}(., v_l)\right), \qquad \forall v_j \in \Gamma^v$$

$$w_v^{k+1} = P_{\mathcal{D}}\left(\tilde{w}_v\right), \qquad \forall x \in \Omega$$

$$\bar{w}_v^{k+1} = 2w_v^{k+1} - \bar{w}_v^k$$

end while

Finally notice that this process involves relations between all the label pairs through the couples $(q_u(x, u_i), q_v(x, v_j))$, so that it leads to store additional variables λ of size $|\Omega|M^2$ during the projection on $\mathcal{Q}$. Such storage can in practice be avoided as the projection is done point-wise and a parallel computation may decrease this dimension as mentioned in [30, 18]. Nevertheless, if one want to avoid the storage of variables of such dimensions, then ρ (which is also of size $|\Omega|M^2$) must be recomputed at each iteration of the process, which has a huge computational cost.

4.2. General convexification. We now look at another convexification that increases the primal dimension but which allows gathering all the unknowns in a single auxiliary variable.

4.2.1. Convexification with upper level sets. As in [9], we can follow the strategy of section 3.1 and consider the function $\phi(x, u_i, v_j) = H(u(x) - u_i)H(v(x) - v_j)$ to convexify the problem (4.1). Note that since $H(u(x) - u_0) = H(v(x) - v_0) = 1$

we have $\phi(x, u_0, v_j) = H(v(x) - v_j)$ and $\phi(x, u_i, v_0) = H(u(x) - u_i)$. Observing that

$$\begin{aligned}
&\int_\Omega \rho(x, u(x), v(x))dx \\
&= \int_\Omega \sum_{i,j=0}^{M-2} \rho(x, u_i, v_j)(H(u(x) - u_i) - H(u(x) - u_{i+1}))(H(v(x) - v_j) - H(v(x) - v_{j+1}))\, dx \\
&= \sum_{i,j=0}^{M-1} \int_\Omega \rho(x, u_i, v_j)(\phi(x, u_i, v_j) - \phi(x, u_i, v_{j+1}) - \phi(x, u_{i+1}, v_j) + \phi(x, u_{i+1}, v_{j+1}))
\end{aligned} \tag{4.17}$$

$$\begin{aligned}
\int_\Omega |Du(x)| &= \sum_{i=1}^{M-1} \int_\Omega (u_i - u_{i-1}) |D_x H(u(x) - u_i)| dx \\
&= \Delta_u \sum_{i=1}^{M-1} \int_\Omega |D_x \phi(x, u_i, v_0)| dx
\end{aligned} \tag{4.18}$$

$$\int_\Omega |Dv(x)| = \Delta_v \sum_{i=1}^{M-1} \int_\Omega |D_x \phi(x, u_0, v_i)| dx \tag{4.19}$$

we can then consider the functional $J(\phi)$ defined as the sum of the three terms (4.17), (4.18) and (4.19) and look at the convex problem

$$\min_{\phi \in \bar{\mathcal{A}}_2} J(\phi), \tag{4.20}$$

defined on the relaxed convex set of admissible functions:

$$\begin{aligned}
\bar{\mathcal{A}}_2 = \{&\phi : (x, s, t) \in BV(\Sigma, [0, 1]),\ \phi(x, u_0, v_0) = 1, \\
&\phi(x, u_i, v_j) - \phi(x, u_i, v_{j+1}) - \phi(x, u_{i+1}, v_j) + \phi(x, u_{i+1}, v_{j+1}) \geq 0 \\
&\phi(x, u_i, u_j) \geq \phi(x, u_{i+1}, v_j),\ \phi(x, u_i, u_j) \geq \phi(x, u_i, v_{j+1})\, \forall i, j\},
\end{aligned}$$

where $\Sigma = \Omega \times \Gamma^u \times \Gamma^v$.

Hence, one can show that if the function ϕ^* is a binary minimizer of J, then the couple (u^*, v^*) defined as

$$\begin{aligned}
u^*(x) &= u_0 + \Delta_u \sum_{i=1}^{M-1} \phi(x, u_i, v_0) \\
v^*(x) &= v_0 + \Delta_v \sum_{i=1}^{M-1} \phi(x, u_0, v_i)
\end{aligned} \tag{4.21}$$

is a global minimizer of the functional (4.1). However, if ϕ^* is not binary, we can not use relations (4.21) to build a solution of the original non convex problem. This is due to the fact that in the $2D$ case, the characteristic functions of the level sets of ϕ may not belong to the set of admissible functions $\mathcal{A}_2$, defined as:

$$\begin{aligned}
\mathcal{A}_2 = \{&\phi : (x, s, t) \in BV(\Sigma, [0, 1]),\ \phi(x, u_0, v_0) = 1, \\
&\phi(x, u_i, v_j) - \phi(x, u_i, v_{j+1}) - \phi(x, u_{i+1}, v_j) + \phi(x, u_{i+1}, v_{j+1}) \in \{0, 1\} \\
&\phi(x, u_i, u_j) \geq \phi(x, u_{i+1}, v_j),\ \phi(x, u_i, u_j) \geq \phi(x, u_i, v_{j+1})\}.
\end{aligned}$$

This main issue will be discussed in section 5. The projection on the derivative constraints included in the convex set $\mathcal{A}_2$ seems here unrealistic due to the complexity of the iterative Dijkstra algorithm that should be done point-wise, for each $x \in \Omega$. For $D > 1$ multi-label problems, there is no relation of order between labels. As a consequence, the results of [27] can no longer be used and the projection is necessary. We now detail the equivalent probabilistic formulation that involves exact projections.

4.3. The $2D$ probabilistic point of view. A probabilistic formulation of the convexification for the $2D$ problem is now proposed. As in section 3.2, we introduce the variable $w(x, u_i, v_j)$ measuring the probability of assigning the label pair $(u_i, v_j) \in \Gamma^u \times \Gamma^v$ to the pixel $x \in \Omega$. The data term now reads:

$$\min_{u\in\Gamma^u, v\in\Gamma^v} \int_\Omega \rho(x, u(x), v(x))dx = \min_{w\in\mathcal{D}_2} \int_\Omega \sum_{i=0}^{M-1}\sum_{j=0}^{M-1} \rho(x, u_i, v_j)w(x, u_i, v_j)dx,$$

where $\Sigma = \Omega \times \Gamma^u \times \Gamma^v$. The previous relation is valid for binary values of w with a single one that is 1 for each $x \in \Omega$. As before, since such class of functions is not convex, the set of admissible functions is relaxed as

$$\mathcal{D}_2 = \left\{ w : \Sigma \mapsto [0, 1], \text{ s.t. } \sum_{i=0}^{M-1}\sum_{j=0}^{M-1} w(x, u_i, v_j) = 1, \forall x \in \Omega \right\}. \tag{4.22}$$

Again, the Total Variation of u and v can be rewritten in terms of w as:

$$\int_\Omega |Du| = \Delta_u \int_\Omega \sum_{i=1}^{M-1} \left| D_x \sum_{j=1}^{M-1}\sum_{k\geq i} w(x, u_k, v_j) \right| dx$$
$$\int_\Omega |Dv| = \Delta_v \int_\Omega \sum_{j=1}^{M-1} \left| D_x \sum_{i=1}^{M-1}\sum_{k\geq j} w(x, u_i, v_k) \right| dx$$

The functional to minimize is then

$$\begin{aligned} J_w(w) = &\int_\Omega \sum_{i=1}^{M-1}\sum_{j=1}^{M-1} \rho(x, u_i, v_j)w(x, u_i, v_j)dx + \Delta_u \int_\Omega \sum_{i=1}^{M-1} \left| D_x \sum_{j=1}^{M-1}\sum_{k\geq i} w(x, u_k, v_j) \right| dx \\ &+ \Delta_v \int_\Omega \sum_{i=j}^{M-1} \left| D_x \sum_{i=1}^{M-1}\sum_{k\geq j} w(x, u_i, v_k) \right| dx. \end{aligned} \tag{4.23}$$

The problem of minimizing (4.23) w.r.t $w \in \mathcal{D}_2$ is convex and a global minimum can be found with the previously introduced dual optimization methods. The process is summed up in Algorithm 7. The projection on $\mathcal{D}_2$ can also be done point-wise by projecting, for each pixel x the vector of coordinates $w(x, ., .)$ onto a simplex of dimension M^2, which makes the primal projection exact, contrary to the upper level set formulation. Also notice that contrary to the previous convexification of the data term, we here do not have any inner loop inside the algorithm, so that there is no approximation with this formulation.

Algorithm 7 Minimizing the 2D convex energy (4.23)

Initialize $w^0 = \bar{w}^0 = \frac{1}{M^2}$, $\mathbf{z}_u^0 = \mathbf{z}_v^0 = 0$, choose $\sigma, \tau > 0$ such that $\sigma\tau < \frac{1}{8M^3}$

while $||w^k - w^{k-1}|| > \epsilon$ **do**

$$\mathbf{z}_u^{k+1}(.,u_i) = P_{\mathcal{B}}\left(\mathbf{z}_u^k(.,u_i) + \sigma D_x\left(\sum_{j'=0}^{M-1}\sum_{i'\geq i}\bar{w}^k(.,u_{i'},v_{j'})\right)\right), \qquad \forall u_i \in \Gamma^u$$

$$\mathbf{z}_v^{k+1}(.,v_j) = P_{\mathcal{B}}\left(\mathbf{z}_v^k(.,v_j) + \sigma D_x\left(\sum_{i'=0}^{M-1}\sum_{j'\geq j}\bar{w}^k(.,u_{i'},v_{j'})\right)\right), \qquad \forall v_j \in \Gamma^v$$

$$\tilde{w}(.,u_i,v_j) = w_u^k(.,u_i) + \tau\left(\sum_{i'\leq i} D_x^*\mathbf{z}_u^{k+1}(.,u_{i'})\right.$$

$$\left. + \sum_{j'\leq j} D_x^*\mathbf{z}_v^{k+1}(.,v_{j'}) - \rho(.,u_i,v_j)\right), \qquad \forall u_i, v_j \in \Gamma^u \times \Gamma^v$$

$$w^{k+1} = P_{\mathcal{D}}(\tilde{w}), \qquad \forall x \in \Omega$$

$$\bar{w}^{k+1} = 2w^{k+1} - w^k$$

end while

Equivalence between the relaxed problems. Probability and upper level sets formulations are equivalent, as in $1D$. Indeed, by taking $\phi(x,s,t) = H(u(x)-s)H(v(x)-t)$ and by considering the problem $\min_{\phi\in\mathcal{A}_2} J(\phi)$, we find the same kind of relation as in section 3.2:

PROPOSITION 4.3. *There exists a bijection f between the convex sets $\mathcal{A}_2$ and $\mathcal{D}_2$, with $f(\phi)(x,u_i,v_j) = \phi(x,u_i,v_j) - \phi(x,u_i,v_{j+1}) - \phi(x,u_{i+1},v_j) + \phi(x,u_{i+1},v_{j+1})$ and $f^{-1}(w)(x,u_i,v_j) = \sum_{k\geq i}\sum_{l\geq j} w(x,u_k,v_l)$ and we have $J(\phi) = J_w(f(\phi))$.*

An example of such relation between local values of $w(x,.,.)$ and $\phi(x,.,.)$ is given in Figure 4.1. Note that the upper level set approach for $2D$ problems has also been recently proposed in [19], where the constraints associated to $\mathcal{A}_2$ are treated with an additional auxiliary variable instead of using a Dijkstra algorithm, as done in [7] for 1D problems.

0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0.4	0
0	0	0.2	0.2	0	0
0	0	0	0	0	0
0.2	0	0	0	0	0

$w^*(x,.,.)$

$\Leftrightarrow$

0	0	0	0	0	0
0	0	0	0	0	0
0.4	0.4	0.4	0.4	0.4	0
0.8	0.8	0.8	0.6	0.4	0
0.8	0.8	0.8	0.6	0.4	0
1	0.8	0.8	0.6	0.4	0

$\phi^*(x,.,.)$

FIG. 4.1. *On the left: example of values $w^*(x,.,.)$. On the right: corresponding values $\phi^*(x,.,.)$.*

The next proposition is a straightforward consequence of what we discussed so far.

PROPOSITION 4.4. *If w^* is a global minimizer of the functional J_w defined in (4.23) w.r.t $w \in \mathcal{D}_2$, then $\phi^* = f^{-1}(w^*)$ is a global minimzer of problem (4.20).*

The method can easily be extended to $N-D$ by following the previous approach and defining the corresponding convex set $\mathcal{D}_N$. With respect to the non convex approach of section 4, we here gain convexity. On the other hand, the dimension of the state explodes, as $|\Omega|M^N > MN|\Omega|$ for practical problems where $M >> 1$. Moreover, the question remains in dimensions $N > 1$ of getting back to a minimizer of the original problem. This is a tough question, as explained in [12], since there may be no connections between the two minimizers. Notice that this issue is not considered in [30], and that the solution proposed in [9] is not satisfactory as will be explained in the next section.

5. How to get back to the original problem ?. In the 1D case, we know how to get a solution of the original problem from the relaxed one. In 2D, the approach with upper level sets does not always give a solution since upper level sets do not always lie in the domain of the original problem. We discuss this further and we propose strategies to go back to the domain of the original problem. We compare the energies of the original problem obtained by these strategies with some other strategies used for the previous formulations.

5.1. Box functions : where it works. In any dimension N, the layer cake formula is valid for the energy J so that

$$J(\phi^*) = \int_0^1 J(H(\phi^* - \mu))d\mu,$$

In dimension $N = 1$, the characteristic function of the level sets of ϕ also belongs to the set of acceptable solutions $\mathcal{A}_1$. This means, as we have already seen in Theorem 3.2, that for almost every threshold $\mu \in [0, 1]$, $H(\phi^* - \mu)$ is a global solution of the convex and non convex problems. In 2D the layer cake formula is still true if we extend the energy to $BV(\Omega \times \Gamma^u \times \Gamma^v, [0, 1])$, but $H(\phi^* - \mu)$ is not bound to be in $\mathcal{A}_2$ and we can not conclude in general.

Next, following [9], we define the box functions set B_f as the set of functions such that for almost every threshold μ, $H(\phi(x, u_i, v_j)-\mu) = H(u_i-u_{k_{x,\mu}})H(v_j-v_{l_{x,\mu}})$. This definition is motivated by the fact that it is then possible to get back by thresholding to the definition domain of the original problem by choosing $(u(x), v(x)) = (u_{k_{x,\mu}}, v_{l_{x,\mu}})$ and we find a global minimum of the original energy. This is illustrated in Figure 5.1.

0	0	0	0	0	0
0	0	0	0	0	0
0.4	0.4	0.4	0.4	0.4	0
0.8	0.8	0.8	0.6	0.4	0
0.8	0.8	0.8	0.6	0.4	0
1	0.8	0.8	0.6	0.4	0

$\phi^*(x, ., .)$

0	0	0	0	0	0
0	0	0	0	0	0
1	1	1	1	**1**	0
1	1	1	1	1	0
1	1	1	1	1	0
1	1	1	1	1	0

$\mu = 0.3$

0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
1	1	1	**1**	0	0
1	1	1	1	0	0
1	1	1	1	0	0

$\mu = 0.5$

0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
1	1	**1**	0	0	0
1	1	1	0	0	0
1	1	1	0	0	0

$\mu = 0.7$

FIG. 5.1. *When $\phi^*(x, ., .)$ is a box function (on the left), then any level set $\mathbf{1}_{\phi^*(x,.,.)>\mu}$ is also a box function (examples are given for thresholds $\mu \in \{0.3, 0.5, 0.7\}$).*

Unfortunately, for general problems, we may have that ϕ^* is not a box function, so that recovering (u, v) involves ambiguities. A sketch of such ambiguity is given in Figure 5.2. Moreover, as the box functions set is not convex, it is difficult to impose such a constraint in practice. Nevertheless, we show in Figure 5.3 the percentage of pixels where we do not find a box function after thresholding in an optical flow estimation. This shows that for real applications, where few global minima of the non convex energy are expected, we can expect recovering interesting solutions of the original problem. Note that this result contradicts the analysis of [19], where the authors observe box functions everywhere in their results.

5.2. Stair functions. As we have seen, we can conclude in the case of box functions, but this case is not general enough for our purpose. Let us explain what happens in dimension 2, if we threshold $\phi_\mu = \mathbf{1}_{\phi^* \geq \mu}$, then $\phi_\mu(x, s, t) \in \{0, 1\}$, but $\phi_\mu(x, u_i, v_j) - \phi_\mu(x, u_i, v_{j+1}) - \phi_\mu(x, u_{i+1}, v_j) + \phi_\mu(x, u_{i+1}, v_{j+1}) \in \{-1, 0, 1\}$. With

0	0	0	0	0	0
0	0	0	0	0	0
0.4	0.4	0.4	0.4	0	0
0.8	0.4	0.4	0.4	0	0
0.8	0.8	0.4	0.4	0	0
1	0.8	0.8	0.6	0	0

$\phi^*(x,.,.)$

0	0	0	0	0	0
0	0	0	0	0	0
1	1	1	**1**	0	0
1	1	1	1	0	0
1	1	1	1	0	0
1	1	1	1	0	0

$\mu = 0.3$

0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
1	0	0	0	0	0
1	**1**	0	0	0	0
1	*1*	1	**1**	0	0

$\mu = 0.5$

0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
1	0	0	0	0	0
1	**1**	0	0	0	0
1	*1*	**1**	0	0	0

$\mu = 0.7$

FIG. 5.2. *When $\phi^*(x,.,.)$ is not a box function (on the left), then its level sets $\mathbf{1}_{\phi^*(x,.,.)>\mu}$ have no reason to be box functions. Examples are given for thresholds $\mu \in \{0.3, 0.5, 0.7\}$. A box function is obtained with $\mu = 0.3$ but stairs functions are obtained with the other thresholds.*

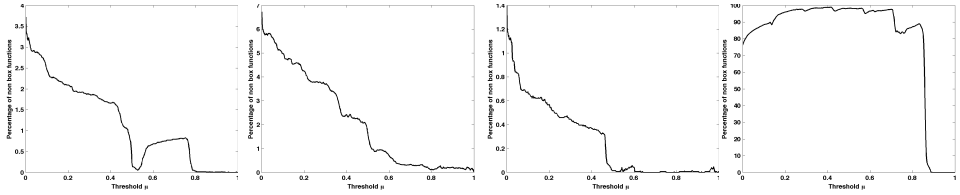


FIG. 5.3. *For any threshold $\mu \in]0,1[$, we plot the percentage of pixels where the function $H(\phi^* - \mu)$ is not a box function. These estimations have been done on 4 different data configurations, the last one being with random data, whereas the other correspond to realistic, where few global minima are expected.*

respect to $w_\mu = \partial_{st}\phi_\mu$, we will have $\sum_s \sum_t w_\mu(x,s,t) = 1$, but $w_\mu(x,s,t) \in \{-1,0,1\}$, which is not acceptable for a distribution.

Let us now assume that at a point x, $H(\phi^* - \mu)$ is not a box function, then $H(\phi^* - \mu)$ is necessarily a binary "stair function" defined in the set S_f:

$$S_f = \{\phi(s,t) : \phi(x,y) \geq \phi(s,t)\, \forall x < s,\, y < t)\}.$$

A sketch of binary stair functions is given in Figure 5.2. The trouble is that stair functions are not necessary box functions, so that the characteristic function of the level sets of ϕ^* do not necessary belong to the set of acceptable solutions $\mathcal{A}_2$.

This explains the reasoning of [9]. The authors advocate the threshold $\mu = 1$ that always lead to a binary box function. The trouble is that such a thresholding is not the best choice as we will see. Another solution to necessarily obtain a box function is to select one (if not unique) of the highest values of $w^*(x,.,.)$ for each $x \in \Omega$, but we will see that that such strategy is not optimal in practice.

5.3. How to choose a box function to approximate a solution of the original problem. In view of the discussion above, a solution to estimate u and v is to consider different thresholdings of ϕ^* that lead to box functions. Let us consider the box function $\phi^b_{\mu,\nu} = H(\phi^*(x,s,v_0) - \mu)H(\phi^*(x,u_0,t) - \nu)$

Regularization term case. We have that

$$\begin{aligned}
&\int_\Omega \sum_{i=1}^{M-1} (u_i - u_{i-1}) |D_x \phi^b_{\mu,\nu}(x,u_i,v_0)| \\
&= \int_\Omega \sum_{i=1}^{M-1} (u_i - u_{i-1}) |D_x H(\phi^*(x,u_i,v_0) - \mu) H(\phi^*(x,u_0,v_0) - \nu)| \\
&= \int_\Omega \sum_{i=1}^{M-1} (u_i - u_{i-1}) |D_x H(\phi^*(x,u_i,v_0) - \mu)|,
\end{aligned}$$

and

$$\int_\Omega \sum_{i=1}^{M-1}(v_i - v_{i-1})|D_x\phi^b_{\mu,\nu}(x,u_0,v_i)| = \int_\Omega \sum_{i=1}^{M-1}(v_i - v_{i-1})|D_xH(\phi^*(x,u_0,v_i)-\nu)|.$$

Denoting as J_{TV}, the total variation part of the energy J, thanks to the coarea formula we therefore see that

$$\begin{aligned}
&\int_0^1\int_0^1 J_{TV}(\phi^b_{\mu,\nu})\,d\nu d\mu\\
=&\int_0^1\int_0^1\left(\int_\Omega\sum_{i=1}^{M-1}\Delta_{u_i}|D_x\phi^b_{\mu,\nu}(x,u_i,v_0)| + \int_\Omega\sum_{i=1}^{M-1}\Delta_{v_i}|D_x\phi^b_{\mu,\nu}(x,u_0,v_i)|\right)d\nu d\mu\\
=&\int_0^1\int_0^1\left(\int_\Omega\sum_{i=1}^{M-1}\Delta_{u_i}|D_xH(\phi^*(x,u_i,v_0)-\mu)| + \int_\Omega\sum_{i=1}^{M-1}\Delta_{v_i}|D_xH(\phi^*(x,u_0,v_i)-\nu)|\right)d\nu d\mu\\
=&\int_\Omega\sum_{i=1}^{M-1}\Delta_{u_i}|D_x\phi^*(x,u_i,v_0)| + \int_\Omega\sum_{i=1}^{M-1}\Delta_{v_i}|D_x\phi^*(x,u_0,v_i)|\\
=&J_{TV}(\phi^*)
\end{aligned}$$

Data term case. For the data term J_{Data}, we need $\int_0^1\int_0^1\phi^b_{\mu,\nu}d\nu d\mu = \phi^*$, but

$$\int_0^1\int_0^1\phi^b_{\mu,\nu}d\nu d\mu = \int_0^1\int_0^1 H(\phi^*(x,s,v_0)-\mu)H(\phi^*(x,u_0,t)-\nu)d\nu d\mu = \phi^*(x,s,v_0)\phi^*(x,u_0,t).$$

where $\Delta_{u_i} = u_i - u_{i-1}$ and $\Delta_{v_i} = v_i - v_{i-1}$. So if $\phi^*(x,s,v_0)\phi^*(x,u_0,t) = \phi^*(x,s,t)$ or, in terms of w, if u and v are independent, then we can find a global solution of the original problem for almost any threshold (μ,ν). If this is not the case, coming back to the upper level sets $\phi^*_\mu = H(\phi^*-\mu)$, one knows that the points (s,t) such that $\phi^*_\mu(x,u_i,v_j)-\phi^*_\mu(x,u_i,v_{j+1})-\phi^*_\mu(x,u_{i+1},v_j)+\phi^*_\mu(x,u_{i+1},v_{j+1}) = 1$ should correspond to high probabilities $w(x,s,t)$, one could then observe that choosing one of these couples to build a box function should be better for the data term J_{Data} of the energy. We denote by (s^i,t^i) these $n_x \geq 1$ points. For each point x and each extremal point (s^i,t^i), we can therefore define a box function $\phi^b_i(x,s,t) = H(s^i-s)H(t^i-t)$. The two box functions associated to the particular extremal points (s^1,t^1) and (s^{n_x},t^{n_x}) are then represented as ϕ^b_1 and ϕ^b_n.

With these notations, the box function $\phi^b_{\mu,\mu}$ is the smallest one to include the binary stair function $H(\phi^*-\mu)$. The coordinates of such box function are represented with the couple of point (s^*,t^*). An illustration of this is given in Figure 5.4 to explicit all these cases. Hence, setting $J_{TV} = J_u + J_v$, we could expect that the different thresholdings of the estimated global solution ϕ^* have the following properties:

- $J_{TV}(\phi^b_{\mu,\mu}) = J_{TV}(\phi^*_\mu) \approx J_{TV}(\phi^*)$ and there is no prior on $J_{Data}(\phi^b_{\mu,\mu})$
- $J_u(\phi^b_1) = J_u(\phi^*_\mu) \approx J_u(\phi^*)$, $J_{Data}(\phi^b_1) \approx J_{Data}(\phi^*)$ and there is no prior on $J_u(\phi^b_1)$
- $J_v(\phi^b_n) = J_v(\phi^*_\mu) \approx J_v(\phi^*)$, $J_{Data}(\phi^b_n) \approx J_{Data}(\phi^*)$ and there is no prior on $J_v(\phi^b_n)$

The energies for the three previously mentioned particular box functions derived from $H(\phi^*-\mu)$ are illustrated in Figure 5.5. These results demonstrate numerically what has been detailed before, apart from the fact that the data term is increasing for large $\mu > 0.5$. This is due to the fact that in the present experiment, the estimated $w(x,.,.)$ appeared to be either unimodal or bimodal for almost any $x \in \Omega$.

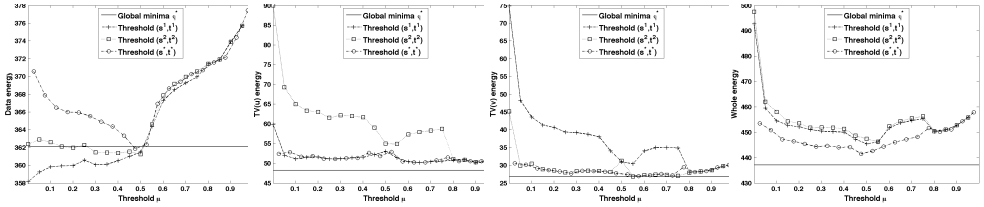
It can be noticed that the first thresholding choice (where the regularization term is controlled by selecting a box function with the pairs (s^*,t^*)) seems to lead to good

0	0	0	0	0	0	0
0	0	0	0	0	0	0
(s^4,t^4) **1**	0	0	0	(s^*,t^*) **0**	0	0
1	(s^3,t^3) **1**	0	0	0	0	0
1	1	1	(s^2,t^2) **1**	0	0	0
1	1	1	1	(s^1,t^1) **1**	0	0
1	1	1	1	1	0	0

FIG. 5.4. *The thresholded solution $H(\phi^* - \mu)$ is given, and the different possible natural binary box functions are shown (defined with the pairs of labels (s,t)). We here have $n_x = 4$.*

results in all the cases. We will therefore consider such a thresholding approach in the following. The best threshold is around $\mu = 0.5$ and we will consider this value from now on. This experiment also shows that the strategy of [9] is not optimal, as the threshold $\mu = 1$ leads to a worse energy.

Remark. Recalling the equivalence between w and ϕ described Figure 4.1, it is interesting to see that box functions will be found anytime as long as we can define, for every $x \in \Omega$ a non decreasing path (s^k, t^k) with $s^{k+1} \geq s^k$ and $t^{k+1} \geq t^k$ that crosses all the non null weights $w(x,s,t)$. Naturally, if we can define for every $x \in \Omega$ a path (s^k, t^k) with $s^{k+1} \leq s^k$ and $t^{k+1} \geq t^k$, then we would have box functions by taking the upper level sets of u and the lower level sets of v. Having such the same configuration of weights for all pixels has nevertheless no reason to occur in general.



(a) (b) (c) (d)

FIG. 5.5. *The energy of the computed global minima ϕ^* is compared to the energies of the different thresholding approaches to obtain box functions. (a) Data term. (b) TV(u). (c) TV(v). (d) Complete energies ((a)+(b)+(c)).*

5.4. Comparison with the other approaches. With respect to the non convex approach [25] (which is a lot faster), the solutions obtained with the proposed convex approach always have smaller energies. We also reach the same global energy minima than the ones obtained with the convexification of the data term in [30]. This is illustrated in Table 5.1, where the energies estimated by the convex and non convex approaches are presented for a real optical flow experiment. The given estimations correspond to a stopping threshold of $\epsilon = 10^{-6}$ in the previously given algorithms. As expected, the solutions obtained with the convex approaches are independent from the initialization. Even if there exist no relations between the non binary global minima of the convex models and the original problems, the thresholding procedure is able to recover solutions of good quality in terms of energy, both with [30] and the proposed model. It is important to notice that the non convex approach with the joint minimization algorithm can produce better estimations of the original problem.

As suggested in [30], we also show the energies obtained by keeping the most probable value for every $x \in \Omega$, that leads to bad estimations only for the proposed convex approach. We also computed the energy associated with the weighted mean value, but do not present them as it gives worse results in practice.

	Relaxed energy: minima estimation	Original energy: best threshold	Original energy: highest probability
Uniform initialization			
Non convex [25]	438.67	441.06	441.58
Non convex joint	432.25	435.48	**435.84**
Data convexified [30]	428.86	435.53	439.19
Proposed convex	428.87	**435.19**	465.23
Best label initialization			
Non convex [25]	461.92	464.89	466.32
Non convex joint	431.57	**434.03**	**434.35**
Data convexified [30]	428.86	435.51	439.35
Proposed convex	428.87	435.23	464.85
Random initialization			
Non convex [25]	602.05	607.72	610.14
Non convex joint	432.01	435.25	**435.44**
Data convexified [30]	428.87	435.63	438.16
Proposed convex	428.87	**435.18**	466.18

TABLE 5.1

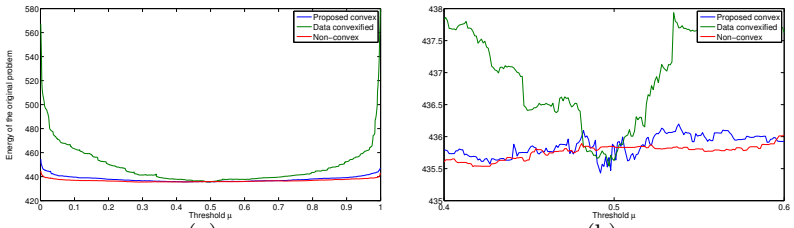
Comparison for a real example. The energies of the minima of the relaxed problems obtained with the different methods and different initializations (uniform, best label without regularization and random) are given in the first column. The original energy is then computed for the solution built by thresholding the corresponding upper level set function (second column) or by keeping the label with the highest probability (last column). The convex approaches lead to the same global minima for any initialization. The non convex approach allows recovering solutions of the original problem whose energies are close to the one of the local solution estimated for the relaxed problem. Note that the new algorithm for the non convex model performs very well.

We show in Figure 5.6 that the non convex method and the proposed convex method give reasonable estimations (in terms of energy) for any threshold, whereas the convex method of [30] is more sensible to thresholding values. The corresponding optical flow estimations are given in Figure 6.5 of next section. Note that in this example realized on a realistic data configuration, we expect the modeling to be well suited, so that the original non convex problem should have few global minima. This explains why the recovered solutions give satisfactory energies.

These examples have been obtained by considering $M_u = 8$ and $M_v = 7$ labels. One can observe that the required memory needed by each approach is:

- Non convex [25]: $3(M_u + M_v)|\Omega|$, as ρ_u, w_u, and $\mathbf{z}_u$ are of size $M_u|\Omega|$, while ρ_v, w_v and $\mathbf{z}_v$ are of size $M_v|\Omega|$
- Non convex joint minimization: $((M_u + M_v)^2 + 2(M_u + M_v))|\Omega|$, as the pre-computed inverse matrix is of size $(M_u + M_v)^2$ at each point x, w_u and $\mathbf{z}_u$ are of size $M_u|\Omega|$, while w_v and $\mathbf{z}_v$ are of size $M_v|\Omega|$
- Convex [30]: $(3M_u + 3M_v + 2M_uM_v)|\Omega|$, as w_u, q_u and $\mathbf{z}_u$ are of size $M_u|\Omega|$ and ρ and λ are of size $M_uM_v|\Omega|$.
- Our approach: $(M_u + M_v + 2M_uM_v)|\Omega|$, as $\mathbf{z}_u$ is of size $M_u|\Omega|$ and ρ and w of size $M_uM_v|\Omega|$.

In all cases, the required memory could be reduced by computing the cost variable ρ when needed, but this would drastically increase the computational cost. Notice that an optimized implementation allows one to reduce the memory requirement of [30] but



(a) (b)

FIG. 5.6. *(a) Comparison of thresholded solutions. The convex approach of [30] (in green) gives very bad estimations for thresholds far from* $\mu = 0.5$. *(b) A zoom of the plot is given for threshold in the interval* $[0.4, 0.6]$.

increase the computational cost. At the present day, the non convex approach of [25] (which is a lot faster) and the convex one of [30] are the only ones being able to deal with very high dimensions. Notice that [30] is nevertheless not exact as it requires inner loops (we realized a fixed number of 50 inner loops in our implementation to approximate with accuracy the projection on $\mathcal{Q}$) and it is in practice more sensible to thresholding.

When increasing the number of labels, the original non convex approach of [25] is more frequently stuck in poor local minima. The new conjoint minimization of the non convex model is nevertheless more robust and is able to produce very good estimations of the original problems in all cases. This is illustrated in Table 5.2 which presents the same kind of results for random data ρ. We also observe that the convex approaches allow one to estimate a better minima of the relaxed energy. However, as the data are totally random, there are a lot of global minima. As a consequence, there are no reasons to recover unimodal probabilities. Hence, when thresholding the solutions given by the convex approaches, it leads to very bad estimations for the original non convex problem, whereas the non convex approach still allows recovering acceptable solutions for the original problem.

	Relaxed energy: minima estimation	Original energy: best threshold	Original energy: highest probability
Non convex [25]	6264	6267	6276
Non convex conjoint	6155	6164	6164
Data convexified [30]	5888	6735	7059
Proposed convex	5888	6725	7308

TABLE 5.2

The energies of the minima of the relaxed problems obtained with the different methods for random data ρ *are given in the first column. The original energy is then computed for the solution built by thresholding the corresponding upper level set function (second column) or by keeping the label with the highest probability (last column).*

We can therefore conclude that:

- The level sets of the solution ϕ^* obtained with the non convex approach have almost the same energy than ϕ^*, but it is not the case for the solution obtained with the convex methods
- The proposed thresholding method associated to our convex method seems good enough for real applications (since the energy is not so far after thresholding to a box function)
- Equivalent solutions are recovered with the model of [30] only with thresholds of value close to 0.5 or keeping the highest probability.
- The improved non convex approach gives quite good local minima with any initialization. This validates experimentally the work of [25].

6. Applications. In this section, we present some numerical results obtained for multi-label problems involving up to $N = 3$ dimensions.

6.1. The $1D$ case.

6.1.1. Color based segmentation. Given one grayscale image I, we look at the segmentation problem

$$J_1(u) = \lambda \int_\Omega |Du(x)| + \int_\Omega |I(x) - u(x)|dx. \tag{6.1}$$

With u taking its values in $\Gamma^u = \{0 < 1/(M-1) \cdots < 1\}$. This corresponds to take $\rho(x, u(x)) = |I(x) - u(x)|$ and we apply Algorithm 3. This is a simple case as the data term ρ is here convex w.r.t u. The result obtained for $\lambda = 10$ and $M = 9$ is presented in Figure 6.1 and compared to the method of [28] (Algorithm 2). The obtained segmentations are here equivalent.

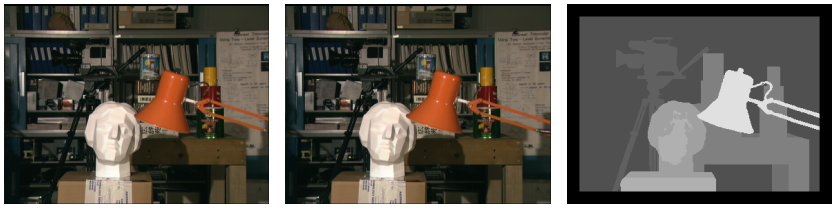
I Our method [28]

FIG. 6.1. *Comparison of the results with [28] with the same parameters $M = 9$ and $\lambda = 4$. Both methods are equivalent.*

6.1.2. Disparity. Given two rectified images I_1 and I_2, (presented in Figure 6.2), we look at the disparity estimation problem

$$J_1(u) = \lambda \int_\Omega |Du(x)| + \int_\Omega |I_1(x) - I_2(x + u(x)|dx, \tag{6.2}$$

with u taking its values in $\Gamma^u = \{4 = u_0 < \cdots < u_{M-1} = 16\}$ and $\lambda = 7$. The data term is non convex as we have $\rho(x, u(x)) = |I_1(x) - I_2(x + u(x))|$. The result obtained with Algorithm 3 is presented in Figure 6.3 and compared with the method of [28] obtained with Algorithm 2, Figure 6.3 shows that both methods are equivalent.



I_1 I_2 u_{truth}

FIG. 6.2. *Stereo data and disparity ground truth.*

6.2. $2D$ optical flow. The optical flow between the images shown in 6.4 has been computed with various approaches: Non convex with joint minimization (Algorithm 5), Data convexified [30] (Algorithm 6) and proposed convex (Algorithm 7). The

Our method [28]

FIG. 6.3. *Comparison of the method with [28] with the same parameters $M = 13$ and $\lambda = 7$. We recover exactlty the same dispartity maps.*

results are presented in Figure 6.5. It illustrates that the non convex approach leads to acceptable estimations with any initialization, while being more robust to the thresholding step.

I_1 I_2 $(u, v)_{truth}$

FIG. 6.4. *Images I_1 and I_2 of the RubberWhale sequence. The optical flow ground truth is presented, the color representing the direction of the flow, while the intensity is related to the norm of the vector.*

The convex approaches are more sensible to thresholding, namely the one of [30]. Notice that the proposed approach allows to recover a thresholded solution with the lower energy, as illustrated in Table 5.1.

6.3. Experiments for $N = 3$. We now show some experiments involving higher dimensions. The following results have been obtained by extending the Algorithm 7 to dimensions 3.

6.3.1. Optical flow with occlusion mask. In order to enhance the optical flow modeling, one could also consider the simultaneous estimation of the occlusion mask. This amounts to increase with one additional dimension to model the occlusion mask. We then have as data term:

$$\rho(x, u(x), v(x), m(x)) = |I_1(x) - I_2(x + w(x))|(1 - m(x)) + \kappa m(x),$$

where $m(x) \in [0, 1]$ represent the occlusion map and $\kappa \geq 0$ is the occlusion cost. As occlusions form connected areas, the occlusion map can therefore be spatially regularized by minimizing:

$$\beta \int_\Omega |\nabla m|,$$

with $\beta > 0$. An example of estimation with $\kappa = 40/255$, $\beta = 0.005$, $\lambda = 0.04$ and comparisons with the proposed convex method (without occlusion mask) as well as the non convex method are given in Figure 6.6. This illustrates the ability of the method to deal with larger dimensions and simultaneously estimate informations of different nature (flow and mask) in a convex way.

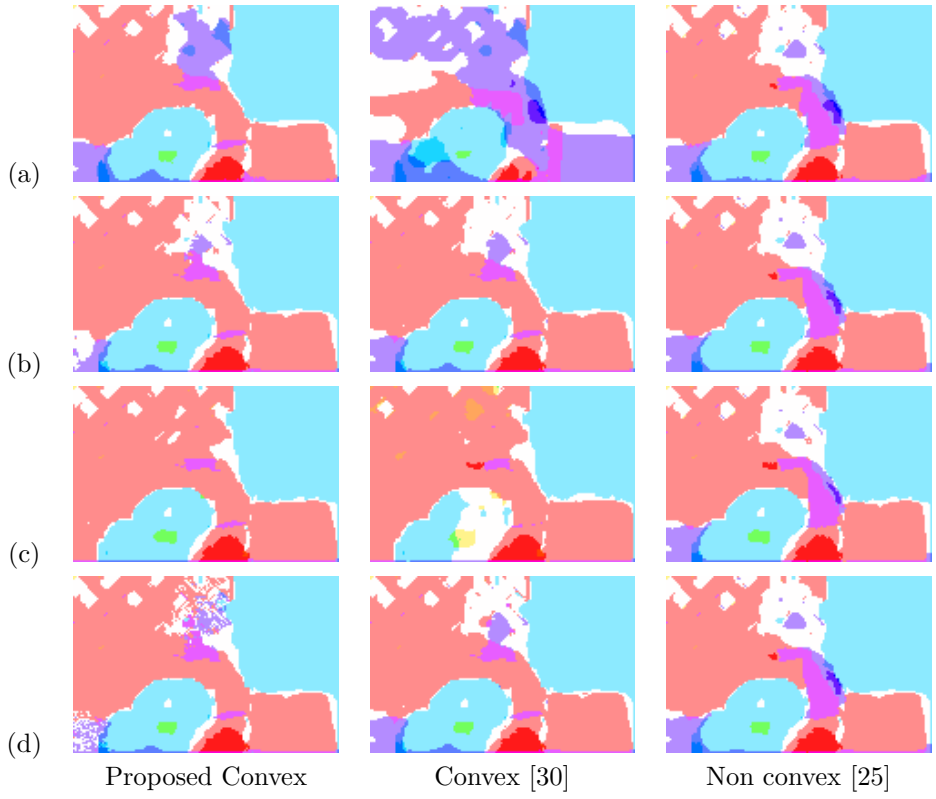


FIG. 6.5. *Comparison of the optical flow between I_1 and I_2 of Figure 6.4 estimated with the proposed convex, convex of [30] and non convex [25] approaches for different thresholds (a: $\mu = 0.1$, b: $\mu = 0.5$, c: $\mu = 0.9$) or highest probability (d). For this simple experiments realized on images of size 146×97 (a 25% rescale of the original size to let the methods converge in reasonable time), we set $\lambda = 0.04$ and only used 8×7 labels in order to visualize the difference between the methods. Note that these results correspond to the Table 5.5, so that the energy of the relaxed solution of [30] is the global minima, but the thresholded solutions are noisy. The proposed convex method allows recovering better solutions of the original problem. The non convex method is less sensible to thresholding.*

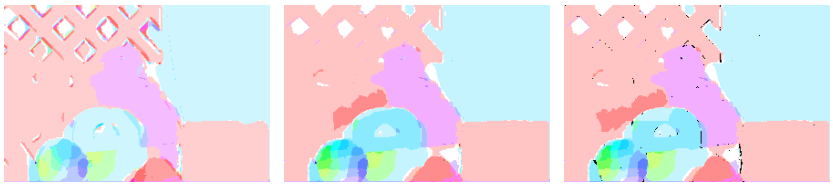


FIG. 6.6. *Solutions obtained on the image RubberWhale of size 292×196 (rescaled at 50% of its original size) with 10×9 labels. The occlusions are represented with black pixels.*

6.3.2. The Chan and Vese model. We consider the Chan and Vese's model:

$$J(u, c_0, c_1) = \int_\Omega ||I - c_0||^2 u dx + \int_\Omega ||I - c_1||^2 (1-u) dx + \lambda \int_\Omega |Du|.$$

Following [9], we convexify the problem by considering the function $w(x, r, s, t) = \delta(u(x) = r)\delta(c_0 = s)\delta(c_1 = t)$ which gives us a data cost:

$$\rho(x, r, s, t) = ||I - s||^2 r + ||I - t||^2 (1 - r).$$

With this formulation, we have to consider maps of constants $c_0(x)$ and $c_1(x)$ and add additional terms: $\kappa \int_\Omega |Dc_0|$ and $\kappa \int_\Omega |Dc_1|$ to the energy, with $\kappa \to \infty$, to ensure recovering constants. Such a modeling is nevertheless not satisfactory. Indeed, if (u, c_0, c_1) is a solution, so is $(1-u, c_1, c_0)$, and also $(1/2, (c_1+c_0)/2, (c_1+c_0)/2)$, by convexity. A constraint like $c_0 < c_1$ should be added. It could be done by adding the penalization $\epsilon(\int_\Omega ||c_0||^2 dx + \int_\Omega ||1-c_1||^2 dx)$, or better by defining the domain of search as $s \leq t$, which corresponds to only treat half of the possibilities.

We applied such a framework as illustrated in Figure 6.7 with a set of possible labels $[0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1]$ which gives us the estimation $c_0 = \frac{1}{4}$ and $c_1 = \frac{3}{4}$. By injecting these estimated constants c_0 and c_1 in a classic TV segmentation model, where c_0 and c_1 are fixed [31], we even exactly recover the same segmentation map with the same energy. It shows that the thresholded solutions are optimal in practice.

FIG. 6.7. *Solution obtained with the full convex model (u^*, c_0^*, c_1^*). The image u_2^* denotes the solution obtained when fixing $c_0 = \frac{1}{4}$ and $c_1 = \frac{3}{4}$. We see that $u^* = u_2^*$, which means that the recovered solution u^* should be optimal.*

7. Conclusion. This paper deals with the computation of global minima of non linear multi-label problems for image processing purposes. We have studied different convex and non-convex relaxations of the non linear problems and we have shown that computing the global minima of a relaxed energy is not sufficient to recover either local or global minima of the original problems. The non-convex approach that estimates a local minima can produce better estimations in practice. These behaviors have been justified theoretically and illustrated experimentally. It therefore appears that in all the presented approaches, there is a duality between being able to compute global minima of relaxed functionals and being able to recover a solution of the original problems. Future work will attempt to deal with such issue.

Acknowledgments. We would like to thank Evgeny Strekalovskiy for giving us technical help for the implementation of [30] and pointing us out the new paper [18]. V. Caselles acknowledges partial support by MICINN project, reference MTM2009-08171, by GRC reference 2009 SGR 773 and by "ICREA Acadèmia" prize for excellence in research funded both by the Generalitat de Catalunya, and by the ERC Advanced Grant INPAINTING (Grant agreement no.: 319899).

REFERENCES

[1] L. AMBROSIO, N. N. FUSCO, AND D. PALLARA, *Functions of Bounded Variation and Free Discontinuity Problems*, Oxford Math. Monogr., The Claredon Press, Oxford University Press, 2000.

[2] G. AUBERT AND J-F. AUJOL, *Optimal partitions, regularized solutions, and application to image classification*, Applicable Analysis, 84 (2005), pp. 15–35.

[3] ———, *A variational approach to remove multiplicative noise*, SIAM Journal on Applied Mathematics, 68 (2008), pp. 925–946.

[4] G. AUBERT AND P. KORNPROBST, *Mathematical Problems in Image Processing*, vol. 147, Springer-Verlag, 2002.

[5] J-F. Aujol, S. Masnou, and S. Ladjal, *Exemplar-based inpainting from a variational point of view*, SIAM Journal on Mathematical Analysis, 42 (2010), pp. 1246–1285.
[6] E. Bae, J. Yuan, and X.-C. Tai, *Global minimization for continuous multiphase partitioning problems using a dual approach*, International Journal of Computer Vision, 92 (2011), pp. 112–129.
[7] A. Baeza, P. Gargallo, V. Caselles, and N. Papadakis, *A narrow band method for the convex formulation of discrete multi-label problems*, Multiscale Modeling & Simulation, 8 (2010), pp. 2048–2078.
[8] J.-M. Ball, *Convexity conditions and existence theorems in nonlinear elasticity*, Archive for Rational Mechanics and Analysis, 63 (1976), pp. 337–403.
[9] E. S. Brown, . F. Chan, and X. Bresson, *Completely convex formulation of the chan-vese image segmentation model*, Int. J. Comput. Vision, 98 (2012), pp. 103–121.
[10] A. Chambolle, *An algorithm for mean curvature motion*, Interfaces and Free Boundaries, 6 (2004), pp. 1–24.
[11] A. Chambolle, D. Cremers, and T. Pock, *A convex approach for computing minimal partitions*, Tech. Report 649, CMAP, 2008.
[12] A. Chambolle, D Cremers, and T. Pock, *A convex approach to minimal partitions*, tech. report, http://hal.archives-ouvertes.fr/hal-00630947, 2011.
[13] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, Journal of Mathematical Imaging and Vision, 40 (2011), pp. 120–145.
[14] T. F. Chan and L. A. Vese, *Active contours without edges*, IEEE Transactions on Image Processing, 10 (2001), pp. 266–277.
[15] Y. Chen and X. Ye, *Projection onto a simplex*, CoRR, abs/1208.4873.
[16] P.L. Combettes and V. Wajs, *Signal recovery by proximal forward-backward splitting*, SIAM J. on Multi. Model. and Simu., 4 (2005).
[17] B. Goldluecke and D. Cremers, *Convex relaxation for multilabel problems with product label spaces*, in European conference on Computer vision (ECCV'10), 2010, pp. 225–238.
[18] B. Goldluecke, E. Strekalovskiy, and D. Cremers, *Tight convex relaxations for vector-valued labeling*, To appear in SIAM Journal on Imaging Science, (2013).
[19] T. Goldstein, X. Bresson, and S. Osher, *Global minimization of markov random fields with applications to optical flow*, Inverse Problems and Imaging, 6 (2012), pp. 623–644.
[20] K. Kolev, M. Klodt, T. Brox, and D. Cremers, *Continuous global optimization in multiview 3d reconstruction*, International Journal of Computer Vision, 84 (2009), pp. 80–96.
[21] P. Maurel, J-F. Aujol, and G. Peyré, *Locally parallel texture modeling*, SIAM Journal on Imaging Sciences, 4 (2011), pp. 413–447.
[22] J.J. Moreau, *Proximité et dualité dans un espace hilbertien*, Bulletin de la S.M.F, 93 (1965).
[23] D. Mumford and J. Shah, *Optimal approximation by piecewise smooth functions and associated variational problems*, Comm. Pure Appl. Math., 42 (1989), pp. 577–685.
[24] M. Nikolova, S. Esedoglu, and T. F. Chan, *Algorithms for finding global minimizers of image segmentation and denoising models*, SIAM Journal on Applied Mathematics, 66 (2006), pp. 1632–1648.
[25] N. Papadakis, A. Baeza, P. Gargallo, and V. Caselles, *Polyconvexification of the multi-label optical flow problem*, in Proc. IEEE Int. Conf. on Image Processing (ICIP'10), 2010.
[26] T. Pock and A. Chambolle, *Diagonal preconditioning for first order primal-dual algorithms in convex optimization*, in IEEE International Conference on Computer Vision (ICCV'11), 2011, pp. 1762 –1769.
[27] T. Pock, D. Cremers, H. Bischof, and A. Chambolle, *Global solutions of variational models with convex regularization*, SIAM Journal on Imaging Sciences, 3 (2010), pp. 1122–1145.
[28] T. Pock, T. Schoenemann, G. Graber, H. Bischof, and D. Cremers, *A convex formulation of continuous multi-label problems*, in Proceedings of the 10th European Conference on Computer Vision (ECCV'08): Part III, 2008, pp. 792–805.
[29] L. I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268.
[30] E. Strekalovskiy, B. Goldluecke, and D. Cremers, *Tight convex relaxations for vector-valued labeling problems*, in IEEE International Conference on Computer Vision (ICCV'11), 2011.
[31] R. Yildizoglu, J-F. Aujol, and N. Papadakis, *Active contours without level sets*, in Proc. IEEE Int. Conf. on Image Processing (ICIP'12), 2012.
[32] C. Zach, T. Pock, and H. Bischof, *A duality based approach for realtime tv-l1 optical flow*, in DAGM conference on Pattern recognition, 2007, pp. 214–223.